



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:

Yin, Muchun

Title:

**A progressively focused qualitative study of teacher thinking in English for academic
purposes (EAP) classroom language assessment**

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

**A PROGRESSIVELY FOCUSED QUALITATIVE STUDY
OF TEACHER THINKING
IN ENGLISH FOR ACADEMIC PURPOSES (EAP)
CLASSROOM LANGUAGE ASSESSMENT**

by Muchun Yin

A dissertation submitted to the University of Bristol in accordance with
the requirements of the degree of Doctor of Education in
the Faculty of Social Sciences and Law, Graduate School of Education

July 2005

Abstract

While a growing number of studies have examined teachers' assessment of students in language classrooms, few have examined the thought processes of teachers as they conduct such assessment, and few have investigated contexts outside primary and secondary education. This dissertation explores the thought processes teachers of English for Academic Purposes (EAP) engage in as they assess students' language ability in the classroom.

The qualitative study reported in this dissertation was progressively focused over two stages. In Stage 1, I broadly explored the cognitions underlying classroom assessment practices, and the sources of these cognitions, through case studies of two instructors teaching an EAP course in a UK university language center; data was collected through classroom observations, interviews, and stimulated recalls. In Stage 2, I examined more closely several issues raised in Stage 1 by conducting three focus groups consisting of EAP teachers.

Overall, the study found important influences upon teachers' assessment thinking, such as teaching approach, classroom parameters, and summative assessments. It also revealed a set of cognitions that teachers drew upon as they assessed students in the classroom, such as the application of assessment principles, stereotyping (in a non-negative sense) often based on ethnicity or nationality, and projection (in which teachers "projected" the student's performance into an imaginary target language use situation and made judgments based on how they thought the hypothetical audience would respond). In addition, the study delineated the uses to which assessment information was put.

Based on these findings, a model of a teacher's "assessment cognition network" is proposed. Other findings relate to the quality of teachers' impressionistic knowledge of students, and the influence of managerial decisions upon teachers' assessment thinking. Recommendations for research, policy, and professional development are also made based upon these findings.

Dedication

This work is dedicated

- ❖ to my father Shoong Yin—a man full of love, generosity, and support—who passed on to be with our heavenly Father during my doctoral studies;
- ❖ to my mother Alma Yin, who has always encouraged me to pursue my academic interests and abilities;
- ❖ and to God—the Father, the Son, and the Holy Spirit—the fear of whom is the beginning of knowledge and by whose grace we all live and move and have our being. *Soli Deo Gloria.*

Knowledge puffs up, but love builds up. The man who thinks he knows something does not yet know as he ought to know. But the man who loves God is known by God.
(1 Corinthians 8:1b-3, NIV)

Acknowledgements

Many people have contributed to this dissertation in ways great and small; I wish to acknowledge some of them here.

I want to express my appreciation for the support of the EdD and PhD students with whom I interacted significantly during my three years at the University of Bristol.

I also want to thank the members of the Centre for Research on Language and Education (CREOLE) for providing such a stimulating intellectual environment and for being a model of a thriving academic community. I especially want to thank the attendees of the CREOLE Research Student Seminar (CRESS), who gave me the opportunity to present parts of this dissertation and provided much useful feedback.

One of the key people for a doctoral student is the supervisor, and I am very thankful to have had Pauline Rea-Dickins as mine; she not only provided support and positive pressure I needed at key points over the past 18 months to move this study forward, but also exemplified over the course of my doctoral studies the kind of formative assessment for learning that is often cited but rarely practiced. She has also served as a role model for me of what an academic should be: a person of expertise, knowledge, and thoughtfulness who also cares very much for students, colleagues, and family, and desires their good.

Deserving of special mention is my family. During the writing of this dissertation, my baby son Joshua Chienyu gave me great motivation with his wonderful smiles and increasing cuteness, which always spurred me on to get my work done so that I could return to the more important and enjoyable task of taking care of him. But the person to whom I owe the greatest debt of gratitude is my wife Yu-lien; she has given me so much support even as she has managed to juggle being both a full-time student and a new mother during the past year. I have also learned much from her about focus, drive, and organization—qualities without which I could not have completed this dissertation. Finally, I am thankful for her love, patience, and encouragement, which have been constant throughout the challenges I have faced during my studies.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original, except where indicated by special reference in the text, and no part of the dissertation has been submitted for any other academic award. Any views expressed in the dissertation are those of the author.

Signed:

A handwritten signature in black ink, appearing to read 'Mueen Zin', written in a cursive style.

Date: 31 August 2005

Table of Contents

Abstract.....	i
Dedication.....	ii
Acknowledgements.....	iii
Author’s Declaration.....	iv
Table of Contents.....	v
 Chapter 1 Introduction.....	 1
1.1 Introduction.....	1
1.2 Classroom assessment and teacher thinking—setting the problem.....	1
1.3 My personal interest.....	3
1.4 Overall research objectives.....	4
1.5 Overview of the research design.....	4
1.6 Overview of the dissertation and summary of chapters.....	6
1.7 Summary.....	6
 Chapter 2 Literature Review.....	 7
2.1 Introduction.....	7
2.2 The construct of classroom assessment.....	7
2.2.1 Dimensions of classroom assessment.....	7
2.2.2 Definitions of classroom assessment and classroom assessment practices	10
2.3 Insights from teacher thinking research.....	12
2.3.1 Beliefs and principles.....	12
2.3.2 Pre/post-active versus interactive thinking.....	15
2.3.3 Routines.....	15
2.3.4 Influences upon cognition.....	16
2.4 A review of research on teacher thinking in relation to classroom assessment.....	17
2.4.1 Models of the assessment process.....	21
2.4.2 Implicit constructs.....	24
2.4.3 Teacher assessment “profiles”.....	25
2.4.4 Contextual influences upon teacher thinking in assessment.....	29
2.4.5 A model of teachers’ classroom assessment decision-making.....	31
2.5 Summary of relevant insights and gaps in knowledge.....	32
2.6 Summary.....	34
 Chapter 3 Methodology: Overview & Stage 1.....	 35
3.1 Introduction.....	35
3.2 Philosophical position.....	35
3.3 Principles guiding the overall research design.....	37
3.4 Rationale for design and conduct of Stage 1 data collection.....	40
3.4.1 Stage 1 research questions.....	40

3.4.2 Case study strategy	40
3.4.3 Data collection methods.....	41
I. Non-participant observation	41
II. One-to-one interviewing	44
III. Stimulated recall	44
3.4.4 Sampling and gaining access	54
3.5 Conduct of Stage 1 analysis.....	57
3.5.1 Summary of Stage 1 analysis.....	57
3.5.2 Generating findings.....	58
3.5.3 Confirming findings.....	63
3.5.4 Concluding comments about Stage 1 analysis.....	65
3.6 Ethical issues arising in Stage 1	65
3.7 Summary	68
 Chapter 4 Stage 1 Findings.....	 69
4.1 Introduction.....	69
4.2 Background to the cases	69
4.3 RQ1: What are teachers' definitions of assessment?.....	73
4.4 RQ2: What are the classroom assessment practices of EAP teachers?	74
4.5 RQ3: What cognitions underlie these practices?	77
4.5.1 Strategic cognitions.....	78
I. Teaching approach and beliefs about language learning.....	78
II. Class parameters.....	79
III. Course syllabus and summative assessments	80
4.5.2 Interactive cognitions.....	82
I. Assessment principles	82
II. Constructs applied in interactive assessment	85
III. Stereotyping	86
IV. Projection.....	88
V. Mental portraits of students	89
VI. Assessments not related to language	90
4.5.3 Uses of assessment information.....	90
4.5.4 An integrated illustration of assessment practices, cognitions, and use ...	93
4.6 RQ4: What are the sources of these cognitions?	95
4.7 General discussion and summary of findings	98
4.8 Issues and research questions for further research.....	102
4.8.1 Quality of teachers' impressionistic knowledge about students	102
4.8.2 The impact of managerialism	103
4.8.3 The use of stereotypes and projection.....	108
4.9 Summary	108
 Chapter 5 Methodology: Stage 2	 110
5.1 Introduction.....	110
5.2 Rationale for design and conduct of Stage 2 data collection.....	111

5.2.1 Stage 2 research questions	111
Issue 1: the quality of teachers' impressionistic knowledge of students	111
Issue 2: the effect of increasing managerialism on assessment thinking.....	111
Issue 3: the use of stereotypes and projection during interactive assessment	111
5.2.2 Deciding upon an interviewing strategy	111
5.2.3 Data collection method	114
The Semi-structured Focus Group.....	114
5.2.4 Piloting the focus group technique	115
5.2.5 Arrangement and conduct of focus groups	116
5.3 Conduct of Analysis.....	118
5.3.1 Summary and description of Stage 2 analysis	118
5.3.2 Issues related to analysis.....	119
5.4 An ethical issue arising in Stage 2	121
5.5 Summary	121
Chapter 6 Stage 2 Findings.....	122
6.1 Introduction.....	122
6.2 Some general remarks.....	122
6.3 RQ5. How do English language teachers think they can increase the quality of their impressionistic knowledge of a student's language abilities?.....	124
6.3.1 Teachers' confidence in their judgments	124
6.3.2 Areas of weakness in teachers' impressions of students	127
6.3.3 Ways of improving teachers' impressions.....	128
6.3.4 Discussion: dealing with blind spots	129
6.4 RQ6. How do teachers who have experienced managerial changes like a more explicit syllabus and more explicit summative assessments feel such changes have affected their assessment practices?	130
6.5 RQ7. Do teachers use stereotypes and projection when they conduct interactive assessment of students?.....	134
6.5.1 Stereotypes.....	134
6.5.2 Projections	137
6.6 Some concluding comments on Stage 2 findings	140
6.7 Summary	142
Chapter 7 Discussion & Conclusion.....	143
7.1 Introduction.....	143
7.2 Summary of the study and its findings	143
7.2.1 Components of the assessment cognition network.....	144
7.2.2 Areas of possible weakness in assessment cognition	146
7.2.3 Organizational influences upon assessment cognitions.....	147
7.3 Limitations of the study	148
7.4 Contribution to knowledge	150
7.4.1 Originality of methodology	150

7.4.2 Teachers' individual assessment practices and cognitions.....	151
7.4.3 The influence of external factors, particularly the organization.....	152
7.4.4 Models of teacher thinking in assessment	152
7.5 Recommendations.....	155
7.5.1 Research.....	155
7.5.2 Policy	157
7.5.3 Professional development.....	159
7.6 Concluding remarks	160
 Bibliography	 161
 Appendix 1 Davison (2004) cline of assessment orientations.....	 169
Appendix 2 Example Stage 1 field notes.....	170
Appendix 3 Classroom action and teacher recall.....	171
Appendix 4 Example contact summary form	174
Appendix 5 Using MAXqda.....	175
Appendix 6 Consent forms used.....	176
Appendix 7 Stage 1 class syllabus and scheme of work.....	181
Appendix 8 Examples of teacher questioning in Stage 1	183
Appendix 9 Interview guide for Stage 2 focus groups	186
Appendix 10 Triangulation of Stage 2 coding.....	188

List of Figures

Figure 2.1 Influences on teacher cognition (from Borg, 2003, p. 82)	16
Figure 2.2 Classroom assessment cycle and processes (from Rea-Dickins, 2001, p. 435)	22
Figure 2.3 Teachers' assessment decision-making (from McMillan & Nash, 2000, p. 11)	31
Figure 3.1 Classification scheme for stimulated recall parameters (from Gass & Mackey, 2000, p. 49, after Faerch & Kasper).....	47
Figure 3.2 Iterations of the assessment cognition network.....	62
Figure 4.1 A continuum of assessment "audiences".....	76
Figure 4.2 Proposed model of an assessment cognition network.....	99
Figure 6.1 Amended model of teachers' assessment cognition network.....	141
Figure 7.1 The assessment cognition network and its sources (same as Figure 6.1)	144
Figure 7.2 Aspects of previous models that are similar to the assessment cognition network model	153

List of Tables

Table 1.1 Structure and timeline of study.....	5
Table 2.1 Literature review grid: Empirical studies of teacher thinking in relation to classroom assessment.....	18
Table 3.1 Overview of research design	38
Table 3.2 Overview of Stage 1 analysis	57
Table 4.1 Classroom assessment practices of teachers of an insessional EAP course, listed in order of prominence/frequency	75
Table 4.2 Teachers' sets of assessment principles (credos).....	83
Table 4.3 An example of a teacher's recall and accompanying cognitions and uses ..	93
Table 4.4 Characteristics of divergent and convergent assessment (from Torrance & Pryor, 1998, p. 153).....	104
Table 4.5 Contrasting aspects of competence and performance models of education (from Broadfoot & Pollard, 2000, p. 20-21).....	106
Table 5.1 Overview of research design (same as Table 3.1)	110
Table 5.2 Overview of Stage 2 analysis	118

Chapter 1 Introduction

1.1 Introduction

This chapter outlines the research problem with which this study is concerned, describes my personal motivation for taking up this work, and sets out the study's overall objectives. It also gives an overview of the study and the rest of this dissertation.

1.2 Classroom assessment and teacher thinking—setting the problem

The surge of research in the area of classroom assessment over the past 10-15 years can be attributed to several factors. One has been the empirical realization that teachers' judgments of students can have important effects upon learning, achievement, and self-perception (Crooks, 1988; Black & Wiliam, 1998). This realization has also been accompanied by theoretical shifts; research in this area, which began in the narrow psychometric orientation of earlier work in testing, has broadened to include a more diverse range of theoretical frameworks (e.g., sociocultural and sociological theories) and conceptualizations of assessment (e.g., assessment as discourse, assessment as a social practice) (see Gipps, 1994; Filer, 2000; McNamara, 2001).

Another significant—and arguably the most powerful—driver of research has been policy changes in many educational contexts around the world, usually in the name of accountability, efficiency, and maintaining or raising standards. One such change has been formalization; whereas teachers have always made informal assessments of students, policymakers are now requiring more formal reporting and stronger, more explicit alignment of teacher assessments with government-mandated curriculum goals (Brindley, 1998, gives an overview in language assessment; for specific examples, Broadfoot & Pollard, 2000, discuss England, and Arkoudis & O'Laughlin, 2004, discuss the Australian state of Victoria.). Much of the classroom assessment research has thus looked at the nature and quality of teachers' assessments under such policies (e.g., Rea-Dickins, 2001; Bachor & Anderson, 1994; McMillan & Nash, 2000). Accompanying this formalization has been increased standardized testing at state or national levels (e.g., assessment accompanying the National Curriculum in the UK and the federal No Child Left Behind mandates in the US); data from these

tests are being used to make decisions about LEAs (local educational authorities), schools, and teachers. Alongside these developments has been a growing interest in the social consequences of large-scale testing; one of these consequences is the effects of such testing on classroom teaching and learning (also known as “washback” effects), which an increasing number of studies have examined (e.g., Alderson & Wall, 1993; Bailey, 1996b; Cheng, 2000; Scott, 2005).

While these factors have converged to result in an increase of research into classroom assessment that has explored many important aspects of the phenomenon, I wish to note two significant areas that have received less coverage.

The first is the thoughts that underlie teachers’ assessment of students. Freeman (2002) has called the mental life of teachers “the hidden side of teaching,”¹ and argues that research into this aspect is key to understanding—and improving—teachers’ professional development and practices. However, only some of the research into classroom assessment has considered this cognitive dimension (these are reviewed in 2.4). At the same time, while there is a significant body of research into teacher cognition—meaning the beliefs, subject knowledge, principles, and thought processes that teachers bring to bear upon their work (see reviews in Shavelson & Stern, 1981; Clark & Peterson, 1986; and Borg, 2003)—it, too, has had little to say in regards to assessment. One aim of this study, then, is to explore “the hidden side of classroom assessment.”

The second is education outside primary and secondary schooling. Understandably, classroom assessment research has focused mainly on primary and secondary contexts for obvious reasons: the impact of assessment is arguably greatest and most wide-ranging upon children and youth, and the school system is a central concern of government policy. However, two considerations justify an expansion of the scope of classroom assessment research. First, there is the great amount of teaching, learning, and assessment occurring in educational institutions outside primary and secondary school—increasingly so due to factors such as the globalization of education (especially higher education), mass immigration, and the emphasis on

¹ The phrase is borrowed from Jackson (1968).

education beyond traditional schooling age (“lifelong learning”) to encourage economic and social productivity. Second, these institutions, like their primary and secondary counterparts, are also operating increasingly in an environment of managerialism and accountability (see Deem, 2004, on this phenomenon in higher education)—an environment in which assessment plays fundamental roles. Another aim of this study, then, is to examine classroom assessment in such educational organizations.

1.3 My personal interest

While this study is an attempt to contribute to the larger research agenda looking at classroom assessment, it is also a personal exercise in sense-making. Over the course of my seven-year professional career as a teacher of English language, particularly EAP (English for Academic Purposes), in universities and language programs in the USA and Taiwan, I have found that many of the puzzles, difficulties, contradictions, and dilemmas I have faced in my work have related to assessment. Working on this study has given me the opportunity to revisit and reflect on my experiences and to theorize about my practice.

I am also motivated by a concern for relevant teacher training in assessment. I am not alone in having had meager and fairly irrelevant initial teacher training in assessment—it seems a widespread problem, as the literature will attest (e.g., McMillan, 2003). I do remember sessions on different types of validity and reliability, computerized testing, and “authentic assessment.”² I do not remember any sessions on handling the tensions and challenges of conducting assessment under realistic classroom conditions. This lack of practical relevance may be related to the delivery model of much initial teacher training (Eraut, 1994), in which an academically-derived body of knowledge is taught to teachers-to-be with the expectation that they will deliver such knowledge to the classroom relatively untransformed. However, it is very likely that I will be the instructor on an initial teacher training unit in assessment when I return to the university from which I am currently taking study leave. It is a personal hope that this study may provide the framework for a different, more relevant assessment training curriculum—one that is

² I use scare quotes not as an implied criticism but as a designation only.

rooted in teachers' classroom experiences and that takes into account the issues that teachers actually must consider during their work.

1.4 Overall research objectives

In view of the aforementioned need to research “the hidden side of classroom assessment,” one main objective of this study is to understand the underlying cognitions that teachers engage in when they assess students in the classroom.

In addition, because of the dearth of research on classroom assessment outside the primary and secondary contexts, and because of my own familiarity with teaching EAP, another objective of this study is to explore classroom assessment in the EAP context.

Below, I briefly discuss the design of the study that was developed to accomplish these objectives.

1.5 Overview of the research design

To examine teacher cognition in EAP classroom language assessment, I decided upon a progressively focused qualitative research design. Broadly speaking, the study was qualitative because I was interested in processes, understandings, and contexts, which I believe are not amenable to quantification. Moreover, the study was “progressively focused” over two stages. As this topic has been relatively under-researched, Stage 1 explored some general initial questions about teacher thinking in relation to classroom language assessment. This exploration was expected not only to provide significant answers to those general questions, but also to raise more specific questions, which were then explored in Stage 2. Table 1.1 shows the general structure and timeline of the study.

Table 1.1 Structure and timeline of study

Stage 1 data collection	Jan 2004-April 2004
case studies: 2 instructors teaching EAP course classroom observation (~40 hours) interview & stimulated recall (~9 hours)	
Stage 1 data analysis	Feb 2004-Dec 2004
Stage 2 data collection	March 2005-April 2005
3 focus groups (of 3-5 EAP teachers each) in 3 different institutions (60-90 minutes each)	
Stage 2 data analysis	March 2005-June 2005
Review of literature	Autumn 2003-April 2005, ongoing
Writing of dissertation	Oct 2004-June 2005

As Table 1.1 indicates, Stage 1 comprised two case studies in which two instructors in a UK university language center teaching an EAP course were observed and then questioned about their cognitions via stimulated recall and interviews. This occurred over one university term and involved about 40 hours of classroom observation and nine hours of interviewing and stimulated recall sessions (see Table 3.2 for details of kind and amount of data collected).

Stage 2 involved three focus groups of EAP teachers at three different UK university language programs. Each group consisted of three to five teachers, and each discussion lasted about 90 minutes. This data confirmed several Stage 1 findings and answered other specific questions arising from Stage 1.

The processes of literature review, analysis, and writing were ongoing; as Table 1.1 shows, one or more of these processes were occurring throughout the whole time period from the preparatory months before January 2004 up to the submission of this dissertation in July 2005.

The specific rationale for this research design and details of its development are discussed at length in chapters 3 and 5.

1.6 Overview of the dissertation and summary of chapters

This dissertation presents a record of the study's design, execution, and findings, and it follows the study's structure. I summarize each of the following chapters below.

Chapter 1 gives background and an overview of the entire study.

Chapter 2 provides key definitions of terms used in this study, highlights concepts from the general education literature regarding teacher thinking, and reviews previous empirical investigations of teacher thinking in relation to assessment. It also points out areas in which further research is needed.

Chapter 3 explains general underpinnings of the entire study's methodology, and then elaborates on the methodology of the study's first stage.

Chapter 4 provides findings from Stage 1. It also proposes a model of teacher thinking in relation to assessment based upon the findings, and sets out more specific issues to be explored in Stage 2.

Chapter 5 explains the methodology of Stage 2.

Chapter 6 reports on the findings from Stage 2.

Chapter 7 summarizes the study's findings and explains the study's contribution to the field. It concludes with recommendations, based upon the findings, for research, policy, and professional development.

1.7 Summary

This introductory chapter has discussed the basis of the research problem and my motivation for conducting this study. It has also described the research goals and given an overview of the empirical study that was designed to reach those goals. Finally, it has provided a guide to the rest of this dissertation. The next chapter looks at previous research relevant to the topic of teacher cognition in relation to classroom assessment.

Chapter 2 Literature Review

2.1 Introduction

In this chapter, I first locate my study in the larger “landscape” of classroom assessment research and define the construct of assessment that I use. I then summarize relevant insights from the literature on teacher thinking research, before finally reviewing studies that have looked specifically at teacher thinking in assessment. Throughout these three sections, I draw upon both the general education and language teaching literature, partly because the amount of language teaching research in the area of teacher thinking and classroom assessment is meager and partly because much of the relevant research from general education also seems directly relevant to language teaching. I have two goals in this chapter; the first is to lay out some initial “conceptual bins” (Miles & Huberman, 1994) with which to begin my empirical work. The second is to argue that assessment as a cognitive process is connected to a larger network of thinking and thus involves a wider array of cognitions than may be traditionally conceived.

2.2 The construct of classroom assessment

2.2.1 *Dimensions of classroom assessment*

The construct of classroom or teacher assessment has been conceptualized in a variety of ways (Gipps, 1994; Brookhart, 2004). For heuristic purposes, I see the various conceptualizations as being located along several dimensions:

- the psychometric dimension; research on this dimension has been concerned with the quality of teacher-made tests, grading, etc., according to principles from large-scale, standardized, psychometric testing.
- the cognitive dimension; research into the cognitive aspects of assessment emphasizes the gathering and interpretation by teachers of information about students (cognitive, affective, etc.) in order to make decisions. Less often, it considers how students interpret assessment decisions made by the teacher (e.g., Brookhart, 2001; Ross, Rolheiser, & Hogaboam-Gray, 2002).
- the process-product dimension; many studies have tried to describe the effects of various assessment techniques or methods upon student achievement. (Reviews can be found in Crooks, 1988, and Black & Wiliam, 1998.).

- the sociological dimension: the classroom is seen as a social microcosm, and assessment as a key element that is connected to larger issues both inside and outside the classroom, such as power relations, personal identity, cultural-historical setting, and the reproduction of social difference, ideology, and inequality (e.g., Filer, 2000; Raveaud, 2004).
- the social constructivist dimension; research here looks at how the teacher provides feedback to students in order to “scaffold” or “construct” their learning. This dimension is mainly associated with research into the formative purpose of assessment. (e.g., Tunstall & Gipps, 1996; Torrance & Pryor, 1998, Leung & Mohan, 2004).

Research along each dimension draws upon different theoretical frameworks and different research methodologies. For example, studies of the constructivist dimension often draw upon concepts associated with sociocultural theories of mind and learning (Moss, 2003) and look at discourse co-produced by teachers and students (e.g., Leung & Mohan, 2004). On the other hand, studies of the cognitive aspect see the teacher (or student) as a person who gathers, interprets, and acts upon data; methodologically, such research often looks at the ways teachers gather and use data (e.g., from tests, portfolios, etc.).

It should be reiterated that these dimensions are only broad categories, are not all incommensurable, and often overlap much in assessment studies. For example, a researcher may look at interactions between teacher and student and analyze how the teacher is not only helping the student learn a concept but also “teaching” the student about proper social behavior in the classroom; this would draw on both constructivist and sociological analyses. Another example is when researchers consider how to improve the psychometric properties of teacher-made tests in order to improve assessment decisions.

In this study, I mainly look at the cognitive side of classroom assessment.¹ One reason I have taken this emphasis is because I wish to look at both assessment and teacher thinking, and a cognitive approach allows me to reconcile or bridge these two areas. In a review of language teacher cognition research, (Borg, 2003, p. 91 &

¹ I do, however, draw on sociological concepts in my analysis in chapter 4 of the relationship between organizational policies and teachers’ assessment cognitions and practices.

93) points out two contrasting research perspectives. One centers on teacher decision-making, “which focuses on identifying the antecedents for teachers’ interactive decisions and describing effective decision-making procedures.” while the other centers on personal practical knowledge, which “examines teaching more holistically, taking into account, for example, the role of affective, moral and emotional factors in shaping teachers’ classroom practices.” It seems that the decision-making perspective in teacher thinking research fits well with a cognitive approach to classroom assessment because they share much in common—they both see the teacher as a thinking agent who makes decisions before, during, and after class time based on a variety of information and considerations.² In fact, one could say that in a cognitive sense, they are inseparable; assessment often provides the information (the antecedents) on which decisions are made.

Another reason for looking at cognition is that, while other approaches have much value, I think a cognitive approach best correlates with teachers’ experiences of assessment. A psychometric perspective is clearly too narrow (Teasdale & Leung, 2000) and I believe only applicable at best to classroom tests. Looking at the sociological dimension would go beyond the teacher to look at wider issues such as those noted earlier; while important, I think they are arguably outside the usual concerns of teachers in their day-to-day work. The same could be said for looking at the effects of particular assessment techniques (i.e., a process-product study); teachers are not usually in a position to study those effects in their daily work. Finally, as for looking at the constructivist aspect of assessment, a teacher must make decisions as action in the classroom unfolds, so it seems to me that he or she cannot examine discourse (as research based on some constructivist perspectives require), simply because it has not been completed yet. This does not mean the teacher cannot do discourse analysis after the class, but that seems to be impractical for teachers in most contexts. It also does not mean he or she cannot consider how to provide formative feedback, but that would also be an internal, cognitive process. In addition, a teacher can gain some information about a student but not provide feedback based on and related to that information; by some constructivist definitions

² This is not to ignore the affective, moral and emotional factors; they play a significant part. However, they are not the main focus in this study. For a fascinating example of a study emphasizing affect in relation to assessment, see Stough & Emmer (1998).

(e.g., Sadler, 1989) this would not be assessment, but it seems to me that that information still contributes to a teacher's overall judgment of a student.

Locating my study in the cognitive dimension of classroom assessment enables me to provide two important definitions.

2.2.2 Definitions of classroom assessment and classroom assessment practices

For this study, I define classroom assessment as “the collection, synthesis, and interpretation of information about student language use in classroom activities to aid the teacher in decision-making.”

To arrive at this definition, I have essentially taken the one provided by Airasian (1997, p. 4)—“the collection, synthesis, and interpretation of information to aid the teacher in decision-making”—and amended it with aspects of the definition provided by Leung (2004, p. 20)—“the noticing and gathering of information about student language use in ordinary (non-contrived) classroom activities, and the use of that information to make decisions about language teaching without necessarily quantifying it or using it for reporting purposes.”

The advantage of Airasian's definition is that it leaves the decision undetermined; that is, the information could be used for formative or summative purposes. Another is that it highlights the interpretive nature of assessment (a la Messick, 1989).

However, Airasian is writing for a general education readership (mainly primary and secondary teachers in training) and includes assessment of aspects such as student motivation and interest level; I am mainly interested in language classrooms and in assessment of student language use,³ and so have narrowed his definition with parts of Leung's (who is writing within the field of language teaching). I have not taken Leung's definition wholesale because it reflects his particular interest in formative assessment, while I am interested in assessment for a range of purposes and in how assessment for different purposes may interact in a teacher's thinking (e.g., how

³ Of course a teacher's assessment of language use will be mixed with assessment of interest level and understanding; again, I mainly want to emphasize my focus rather than draw strict boundaries around my object of study

summative assessments impact on formative assessments and vice versa (see Harlen & Deakin-Crick, 2002, for discussion of this in regards to the effects of large-scale testing).

The definition I have set out above is a mental one; an accompanying term that I use in this study is “classroom assessment practices,” which are the observable phenomena that manifest assessment. Such practices have been described in a number of studies (e.g., Mavrommatis, 1997, looked at the classroom assessment practices of Greek primary teachers; Bachor & Anderson, 1994, of primary teachers in British Columbia; and Cheng, Rogers, & Hu, 2004, of ESL/EFL teachers at universities in Canada, Hong Kong, and China), but Airasian (1997) uses two broad categories: paper-and-pencil techniques and observation. Stiggins & Conklin (1992) categorize classroom assessment as occurring via four main kinds of activities: published tests used by the teacher, teacher-made tests, structured performances (in which assessment is planned by the teacher), and spontaneous performance (in which the teacher incidentally assesses as a student does a “pedagogic” activity). As these and other writers point out, the activities vary in formality (i.e., degree of structure and planning) and standardization, and the resulting assessment information can be used for a variety of purposes. In addition, the frequency with which these activities are used will vary with the teacher and the class.

There is a great deal of ambiguity in the literature over whether observation during an activity designed for teaching counts as assessment; many of the studies included in this review do not include such observational assessment, or only mention it in passing as one means of assessment, while not explicitly studying it. However, I have included this within the purview of the study because both in personal experience and empirical research (see 2.4), teachers do gain information about students through this practice.

With definitions of classroom assessment and practices in place, I now wish to turn to the literature on teacher thinking in order to draw out further insights on assessment. Again, let me reiterate that I have not attempted to cover all the research on teacher thinking, but only those concepts I have come across in my reading of the

literature that may have significant relevance to teacher assessment and may be useful in collecting and analyzing data.

2.3 Insights from teacher thinking research

The research into teacher thinking has grown steadily over the past 30 or so years (Freeman, 2002). According to Clark & Peterson (1986, p. 255-256), Jackson's (1968) *Life in Classrooms* was seminal in this regard; the experimental/correlational/process-product approach was paradigmatic in educational research at the time, but Jackson's descriptive study of teachers' thinking underlying behavior "portrayed the full complexity of the teacher's task ... and called the attention of the educational research community to the importance of describing the thinking and planning of teachers as a means to fuller understanding of classroom processes." Teacher thinking research thus takes as a foundational tenet a view of the teacher as a thoughtful professional rather than a script-following technician (Borg, 2003; Clark & Peterson, 1986).

Below, I discuss key concepts from teacher thinking (also known as "teacher cognition" or "teacher decision-making") research and their relevance to assessment thought processes and practices.

2.3.1 *Beliefs and principles*

Surveys of the literature often highlight the profusion of terms used to examine the implicit aspects of teacher thinking (Borg, 2003; Clark & Peterson, 1986; Pajares, 1992; Shavelson & Stern, 1981). A partial list would include: teacher's personal perspective, principles of practice, construct system, practical knowledge, personal practical knowledge, belief-assumptions-knowledge (BAK) systems, and implicit theories. As Clark & Peterson (1986, p. 287) point out, while the terms differ to some degree in meaning, "they hold in common the idea that a teacher's cognitive and other behaviors are guided by and make sense in relation to a personally held system of beliefs, values, and principles."

It follows, then, that while a cognitive definition of assessment like the one that I give in 2.2.2 could be viewed narrowly in a rather technical, computer-like manner, a more appropriate view of assessment would be to see its collection, synthesis, and interpretation aspects as parts of a wider system of thought processes. Therefore, this study aims to explore those aspects of teacher thinking that are closely connected to assessment processes. However, precisely because teacher cognition is conceptually a network or a system, I would argue that it is not possible to extricate thinking related to assessment out of other cognitions; rather I see the task of this study as highlighting the “strands” of assessment thinking from the greater “weaving” of a teacher’s thinking in general.

At this point, I wish to highlight two important aspects of the cognition “system”—beliefs and principles. I consider them here together mainly because researchers in teacher cognition have emphasized the difficulty of differentiating mental concepts (understandably, if they are all parts of an interconnected network of thinking); for example, Grossman, Wilson & Shulman (1989) in Borg (2003) state that distinctions between beliefs, knowledge, theories, etc. are “blurry at best,” and Woods (1996) used the term “BAK system” in his study because he felt he could not conceptually distinguish between beliefs, assumptions, and knowledge. Thus, while I do draw a distinction between beliefs and principles, it must be kept in mind that the difference is mainly for heuristic value; I do not wish to draw hard lines between what is considered to be a “belief” versus a “principle.”

Beliefs, sometimes known as implicit theories, are “reasonably explicit ‘propositions’ about the characteristics of objects or object classes,” (Nisbett and Ross, 1980, in Clark & Peterson, 1986, p. 281) with “objects” including things such as teacher responsibilities, students, subject matter, the curriculum, and the teaching context. These beliefs may often be implicit to the teacher but become explicit as the teacher interacts with a researcher.

Of particular interest to this study are teacher beliefs about language and language learning. Borg (2003) reviews a number of studies that found that teachers had personal, idiosyncratic beliefs about and knowledge of language, language teaching, and language learning—mainly in relation to grammar, reading, and writing. For

example, some teachers strongly believe that a solid grounding in grammar is necessary if students are to be good users of the target language, while other teachers strongly feel otherwise. These beliefs may have an impact on how or what teachers assess; one of this study's aims is to explore what beliefs teachers hold that impinge upon assessment,

I see beliefs as more theoretical, while principles are often based upon underlying beliefs but are more practical and action-oriented in nature (Richards, 1998). Also known by a variety of other labels—e.g., “principles of practice” (Clark & Peterson, 1986), “maxims” (Richards, 1998), “pedagogic principles” (Breen, Hird, Milton, Oliver, & Thwaite, 2001)—the term “principles” refers to broad mental rules that guide a teacher's decisions. For example, Bailey (1996a) found that teachers used principles like “serve the common good” and “teach to the moment” when deciding to go away from their original plans during a lesson. The teachers in Richards (1998) followed maxims such as “maintain active student involvement.”

However, such principles are not rigidly followed. They are flexible and may contradict each other at times. Besides, as Calderhead (1984, p. 91) points out, while “teachers themselves may well have a set of personal beliefs about the nature of teaching and how they should carry out their work [what I am designating “principles”], ... in the process of translating these into action, other factors frequently seem to have a powerful effect upon the outcome.” Thus, as Clark and Peterson (1986, p. 290) conclude, “principles of practice, while useful as general guides for planning, organizing, and teaching in the classroom, are not sufficient by themselves and require artful interpretation, balance, compromise, and, occasionally, intentional violation to serve the experienced teacher well.”

Many of the studies examining teachers' principles have considered general ones guiding classroom conduct and management. Studies of teacher thinking in specific areas (e.g., reading) have also shown that teachers abide by certain principles (Borg, 2003). It is thus reasonable to assume that principles also guide teachers' thinking in assessment. This too will be explored in this study.

2.3.2 *Pre/post-active versus interactive thinking*

As opposed to the aforementioned ambiguity of terms for mental concepts, the distinction that researchers have made between thinking that occurs before and after class time (pre/post-active) and thinking during class time (interactive) has been very robust. Pre/post-active thinking consists, among other things, of planning and evaluating a particular lesson, considering the activities that will be used and how they “flow” together, and allocating time and resources for activities and students. Interactive thinking, on the other hand, involves ongoing decision-making or monitoring in the classroom that is partly shaped by planning but is contingent upon a variety of factors such as student response. In fact, it is very telling that, according to Clark & Peterson (1986), studies of interactive thinking show teachers thinking about students about 40-50 percent of the time, on average.

This distinction is relevant to assessment in two ways. One is that assessment can be seen as occurring pre/post-actively or interactively; the former would be conducted with relatively less time pressure and can involve more evidence of student ability (e.g., looking back at a student’s previous assessment performance, looking over a student’s work more than once), while the latter would be conducted with greater time pressure and less evidence. The other way is that planning for assessment would normally occur pre/post actively. This planning stage is actually widely noted in the assessment literature (e.g., Airasian, 1997; Hall, Webber, Varley, Young, & Dorman, 1997; Rea-Dickins, 2001) and is a key site of teacher cognition. I discuss this further in 2.4 below.

In this study, for ease and clarity of reference, I call pre- and post-active thinking together—because they blend together in practice—as “planning” and I use the terms “interactive cognitions” and “interactive assessment” to refer to the ongoing, “in-flight” (Rea-Dickins, 2003) thinking and assessment that occur in the classroom.

2.3.3 *Routines*

The research literature on teacher thinking also makes prominent the role that routines play: “These routines are the shared, scripted, virtually automated pieces of action that constitute so much of our daily lives [as teachers].” (Berliner in Richards,

1998, p 74) Given that teachers have to process enormous amounts of information in the course of their work, routinization can reduce the cognitive burden. These routines help “remove doubts about what to do next, reduce complexity, and increase predictability.” (Crookes & Arakaki, in Borg, 2003, p. 87)

By implication, thought processes related to assessment would also display some form of routinization. One example can be found in Davison (2004, p. 316). The teachers who participated in the study were given texts and asked to assign grades. One teacher states, “You’re thinking globally first off on first reading, and then you start to apply the criteria [provided in a government framework] and your frame of mind changes according to ‘Do I downgrade?’ or ‘Do I upgrade?’...” This suggests that teachers in the classroom may also follow similar routines when assessing students; however, this remains to be empirically examined.

2.3.4 Influences upon cognition

The teaching thinking research has not only tried to describe the thought processes deployed by teachers in their work, but also attempted to understand the sources of and influences upon them. Borg (2003) provides a framework for understanding some of the major influences (Figure 2.1).

Figure 2.1 Influences on teacher cognition (from Borg, 2003, p. 82)

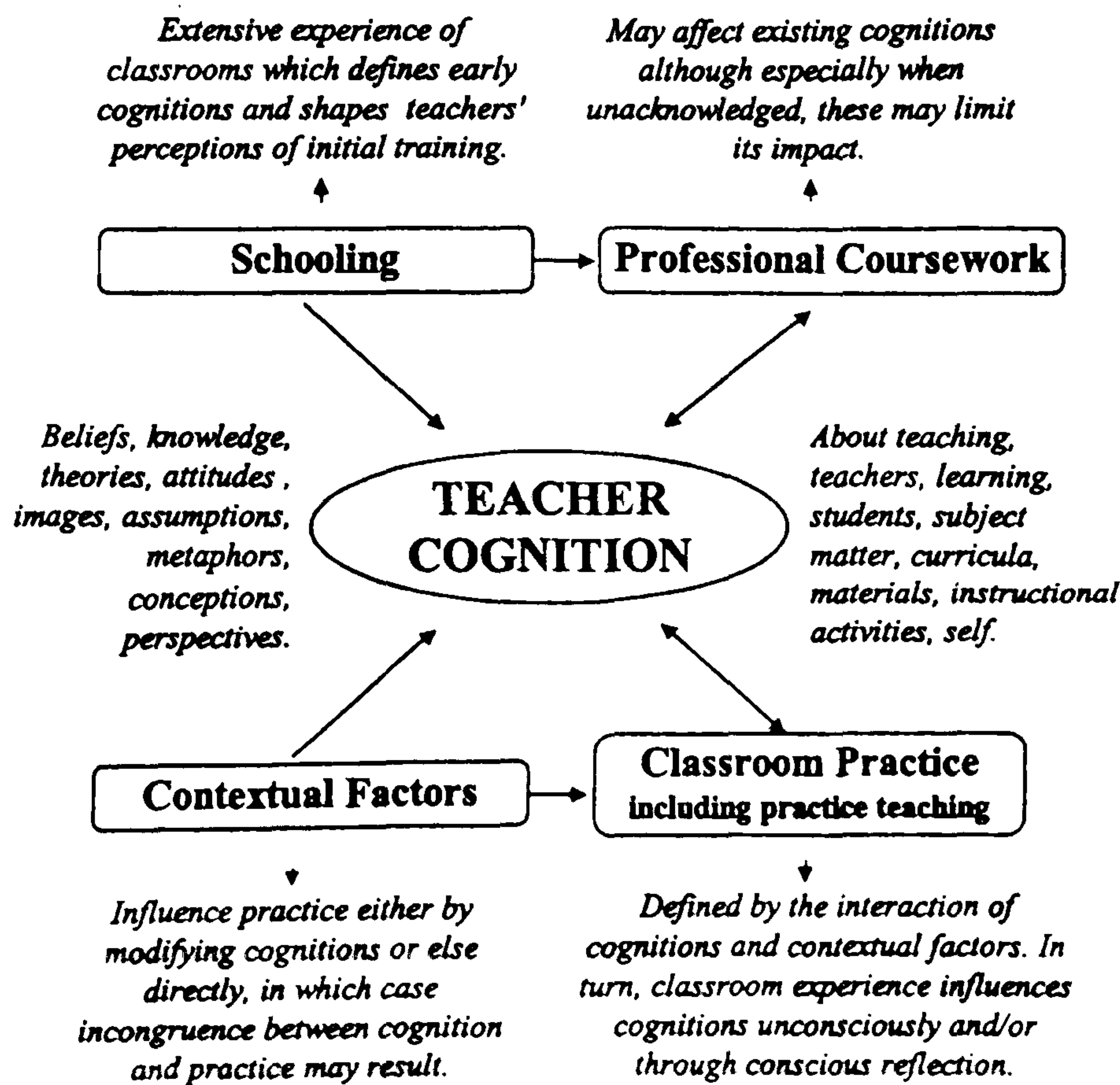


Figure 2.1 shows that the main influences upon teacher cognition have been found to be schooling, contextual factors, professional coursework, and classroom practice, with the latter two being in reciprocal relationship to teacher thinking; that is, a teacher's beliefs, knowledge, etc. both affect and are affected by the teacher's experiences in professional coursework and in the classroom.

If teacher assessment, as one of the cognitive processes that teachers engage in, is connected to a wider network of beliefs, principles, etc., then it follows that teacher assessment also is influenced by a variety of factors. One major goal of this study is to understand what those factors are and how they impinge upon assessment thinking.

In this section, I have highlighted important concepts from teacher cognition research that are likely to have relevance to assessment thinking. In the next section, I review the research done specifically in teacher thinking related to assessment.

2.4 A review of research on teacher thinking in relation to classroom assessment

There have been a number of assessment studies in which teachers' thought processes were a relatively significant topic of study (see Table 2.1). These studies were culled from personal reading of assessment journals, suggestions from my supervisor, and database and webpage searches.⁴

⁴ I searched mainly using the keywords teacher thinking/cognition/decision-making beliefs/knowledge/attitudes and classroom assessment/classroom evaluation/teacher assessment student evaluation; the databases searched included the MLA Bibliography, PsycINFO, ERIC, and ISI Web of Knowledge.

Table 2.1 Literature review grid: Empirical studies of teacher thinking in relation to classroom assessment

* These articles reported portions of significantly larger studies, so the actual number of participants could be larger and often more mixed (e.g., to include students, administrators, etc.), and/or the findings reported in the cited article were part of a larger set of findings.
** Multiple references refer to the same study.

Reference	Research Questions/Foci	Participants	Method(s)	Relevant Findings
Bachor & Anderson (1994)	To describe the classroom-based assessment practices employed in the Province of British Columbia	Over 240 Grades 3/4 and 6/7 teachers from 10 LEAs*	Teachers were interviewed in pairs about views of assessment during policy changes in BC.	Results essentially replicated McCallum et al. (1993) but with some differences.
Cumming (2001; 2003)	To uncover the nature of differences in ESL/EFL teachers' writing assessment practices.	48 highly experienced ESL/EFL writing instructors at universities in Australia, Canada, Hong Kong, Japan, New Zealand, and Thailand*	Instructors were interviewed.	Participants seemed to have either a specific-purpose or a general-purpose orientation in conceptualizing ESL/EFL writing curricula, and this influenced their approaches to assessment.
Davison (2004)	To explore the extent to which senior secondary English language teachers in Melbourne and Hong Kong schools shared common (and stable) beliefs and values about the construct being assessed, with specific reference to written argument, and the extent to which teachers felt their judgments were legitimated and "trusted" in their community.	12 teachers of ESL in each city	Teachers were asked to assess 6 texts according to their "usual" processes and "think aloud." Then in groups of 3 they shared and reflected on their results and processes. A week later, all 12 teachers met as a group to discuss contradictions in practices and attitudes.	Melbourne teachers shared much in common (e.g., using a standard framework) in their processes, while HK teachers were more diverse in their criteria. Melbourne teachers seemed to put more weight on their professional judgments versus external standards, while HK teachers seemed to feel their judgments would not be respected unless attached to external exams. A cline of teachers' assessment orientations is proposed.
Edelenbos & Kubanek-German (2004)	A. What are the instructional behavior patterns shown in class by EFL teachers in primary schools? During which instructional patterns do teachers observe and monitor students' language production? B & C. Primary foreign language teachers' "diagnostic competence"	A. 20 Dutch primary schools (10 from 1987, 10 from 1995), with 24 lessons (1987) and 25 lessons (1995) B. A German primary school teacher C. 3 primary foreign language teachers of 12 German classes, total of 150 students, observed between fall 2000 and summer 2002	A. 1987 classes were videotaped, leading to an observation schedule of general instructional behavior. Teachers also commented on lessons (stimulated recall). An observation schedule of diagnostic behavior resulted from the above two. This was used to code the 50 [sic] lessons. B. Stimulated recall of an EFL lesson C. Ethnographic study of teachers and "diagnostic competence," including interviews, videotaped lessons and field trips, teachers' diaries, teachers' notes on children's assessment papers, field notes, etc.	A. From the 1995 sample, teachers spent 11.3% of class time on diagnostic behavior. These behaviors are described (e.g., "questioning" and "administering and interpreting a test"). B. The teacher's diagnostic behavior and pedagogical responses were connected to lesson planning, children's language acquisition, whole-class assessments, assessment of strong and weak learners, and a subjective theory about language learning. C. Three types of teachers are profiled. Based on these studies, the authors describe 6 levels of "diagnostic competence."

Reference	Research Questions/Foci	Participants	Method(s)	Relevant Findings
Hall & Harding (2002)	To examine extent to which practitioners engaged in a community of assessment practice	Year 2 and 3 teachers and assessment coordinators from 6 schools from 6 LEAs in northern England	Participants were interviewed (some twice in two years, some once) about mechanisms and structures at school level to promote and share assessment criteria with staff, pupils, and parents	Some schools were collaborative while most were individualistic. Also, such collaboration did not extend between schools.
Hall, Webber, Varley, Young, & Dorman (1997)	To describe procedures and manageability of teacher assessment, and the impact of assessment on children's learning and teachers' pedagogical approaches	50 Year 2 teachers in 45 schools in one English LEA.	Teachers were interviewed about assessment.	Practices are described and a developmental model of teacher assessment is proposed.
Leung & Teasdale (1997)	To what extent is there a common set of criteria which primary school teachers (in England) use to make assessments? What are the relationships between the features identified by teachers? To what extent do teachers use different criteria to those contained in the National Curriculum (NC) level descriptors?	A. 12 Year 3 pupils B. 9 primary teachers C. 4 panels of primary school teachers (data from 39 teachers)	A. Pupils were videotaped in primary classes; tape edited into 12 performances. B. Teachers watched performances and were interviewed about them with repertory grid technique to produce constructs. C. Panels watched the 12 performances, rated them according to NC descriptors and constructs from B, differentiated between English-as-Mother-Tongue and EAL pupils, and rated usability of constructs from B.	Teachers had some shared understandings of general criteria to assess speaking and listening, particularly around the construct "native speaker-like"; however, these criteria were not all covered in the NC level descriptors. Also, the teacher-reported constructs did seem to have overlap and redundancy.
Mavrommatis (1997)	To understand Greek primary teachers' views about their classroom assessment practices and how they actually implemented them	A. 372 primary school teachers were surveyed B. 20 of these were observed	Survey and then classroom observation (some for two full days and some for one day)	A 4-phase "assessment episode" is developed and then used to describe and examine assessment practices: evidence collection, interpretation, teacher's response, and impact on pupils.
McCallum, McAlister, Brown, & Gipps (1993)	To monitor implementation of the Teacher Assessment aspect of the National Curriculum, and interpretation and use of results	25 Year 2 teachers from schools in 4 LEAs around England*	Teachers were interviewed and then responded to a list of quotes; analysis of this data led to development of 3 models of teacher assessment, which teachers were then asked to respond to.	3 models of teacher assessment are proposed: Critical Intuitives, who saw the assessment system as a disruption of intuitive practices; Evidence Gatherers, who mainly collected evidence that was evaluated later (they adapted the new assessment system to a degree); Systematic Planners, who "embraced" the new assessment system and integrated it with teaching.

Reference	Research Questions/Foci	Participants	Method(s)	Relevant Findings
McMillan (2003), McMillan & Nash (2000)**	What is the nature of teacher decision making concerning classroom assessment and grading practices? What influences teacher decision making concerning classroom assessment and grading practices? What justification do teachers give for their classroom assessment and grading practices?	24 teachers in the USA (1 5 th grade teacher, 10 English teachers, 13 math teachers from 13 schools in 7 school districts)	Teachers were interviewed about their assessment and grading practices.	6 major themes found: Teacher beliefs and values; Classroom realities; External factors; Decision making rationale; Assessment practices; Grading practices. Under each theme were more specific categories. A model of teachers' assessment and grading practices decision making, incorporating the 6 themes, is also proposed.
Orrell (1995)	On what basis were assessment tasks designed and what qualities did they manifest for supporting appropriate learning for higher education? What frames of reference and strategies did academics use when they assessed their students' written assignments to interpret and to attribute a grade to the assignments? What was the source of theses frames of reference and the strategies that were used?	16 experienced academics from education and nursing in Australia*	Each participant conducted a think-aloud protocol as he/she assessed and graded 3 of his/her students' written assignments. Each participant was then interviewed a month later and responded to a set of contradictory statements about the purposes, criteria, practices, and principles of assessment.	Assessment tasks required complex thinking and were grounded in representative real-world problems. When assessing, participants judged student work according to multi-faceted "reference models" that were based largely upon personal experience and "wisdom of experience"; these and theory were main sources of assessment practices and frames of reference.
Rea-Dickins (2001)	To identify processes of assessment in EAL classrooms; to understand the different identities of assessment	2 language-support teachers and 1 mainstream class teacher in an English primary school*	3 lessons involving assessment were observed and videotaped, and teachers were interviewed about the lessons	An "assessment cycle" is identified, as are strategies deployed at each stage of the cycle. Bureaucratic, pedagogic, and learning identities are also identified.
Realí, Reyes, Martucci, Mizukami, Lima, Tancredi, & Mello (2001)	To investigate Brazilian elementary teachers' decision-making processes when assessing students' written work	25 teachers from the first four grades of one school, plus the principal and pedagogical coordinator	Participants were given works written by 4 students and asked to give a written report of the grades they gave and rationale.	Participants showed 4 assessment patterns, varying on how they judged aspects internal and external to the text, and how secure they were of their judgments.
Samuelowicz & Bain (2002)	To identify academics' orientations to assessment practice	20 academics from 5 disciplines at 3 universities in Australia	Academics were interviewed about specific assessment tasks they used and about general issues related to assessment.	A framework for categorizing academics' orientations to assessment practice in terms of belief dimensions is identified.
Yung (2001; 2002)**	To examine the relationship between teachers' beliefs and their practices under an assessment reform (TAS)	10 biology teachers in Hong Kong	Teachers were observed in class, then interviewed afterwards both about lessons and about wider issues.	Teachers' attitudes towards TAS varied individually in terms of active vs. passive response and the metaphors they used to describe their roles in assessment.

As a cursory examination of the table shows, it is an eclectic collection; the studies cover a variety of topics, locations, participants, subjects, and methods. Notably, I have excluded studies of implicit constructs of raters and effects of high-stakes tests on teaching, as these are specific areas each with a significant body of research: I limit here my discussion of these two topics to their relationship with classroom assessment thinking.

In this section, I discuss the above studies according to a few main topics: models of the assessment process, implicit constructs, teacher assessment profiles, and context effects on teacher thinking. I conclude with a discussion of McMillan & Nash's study, as it incorporates several of these topics in the model it proposes.

2.4.1 Models of the assessment process

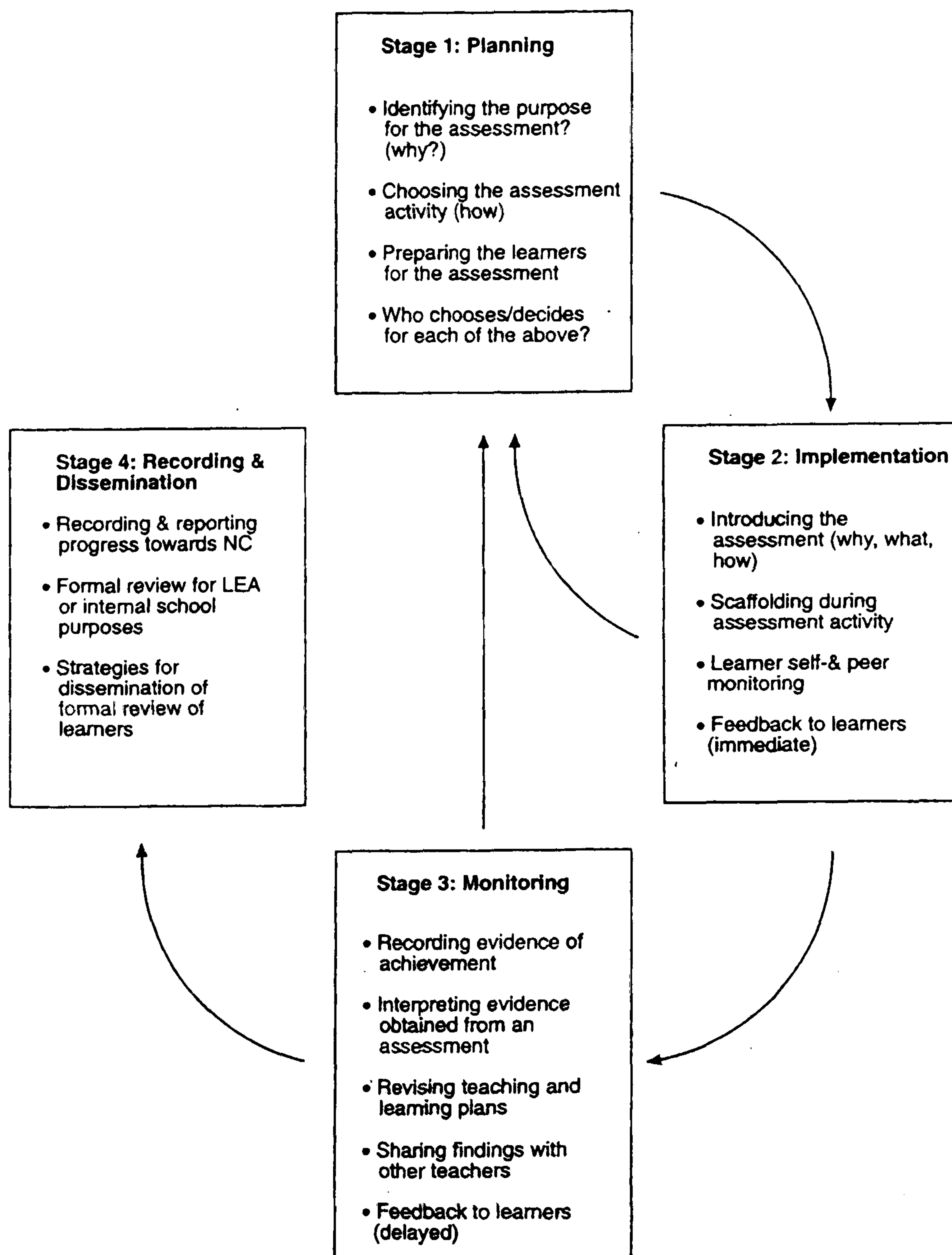
Various models of the classroom assessment process have been put forth. Here, I discuss three studies that have put forth models; these were chosen because they are relatively recent and they are based on strong empirical evidence. The three are discussed in order of increasing scope.

The first is by Mavrommatis (1997). In an analysis of assessment in Greek primary classrooms, he suggests that teacher assessment occurs through a completed assessment activity (which can range in formality)—what he calls an “assessment episode.” Such episodes involve 4 phases:

- 1) Evidence Collection—The teachers he observed used a variety of practices, the most common being observations of academic, behavioral, and social characteristics, and the others being oral questioning, written textbook tasks, and teacher-made tests.
- 2) Interpretation—Teachers had criteria that were somewhat vague and often focused on non-cognitive aspects; they also used norm-referenced and ipsative-referenced judgments.
- 3) Teacher Response—These were often nonverbal, such as a certain look to control students; they also included oral and written comments, marking, and grading.
- 4) Impact on Students—This was not explored as fully as the other phases, but the data showed that students were often concerned with how peers viewed them in light of assessment results.

A model with wider scope is presented in Rea-Dickins (2001), based upon a study of English as an Additional Language (EAL)⁵ learners in English primary schools; she proposes a 4-stage assessment cycle (Figure 2.2).

Figure 2.2 Classroom assessment cycle and processes (from Rea-Dickins, 2001, p. 435)



Briefly, the stages shown in Figure 2.2 are as follows:

- 1) Planning—The teacher makes decisions about what and how to assess, and how to prepare students.
- 2) Implementation—In-class decisions regarding introducing and executing the assessment are made.

⁵ In this study, EAL and ESL (English as a Second Language) are used interchangeably. EFL is used for English as a Foreign Language.

3) Monitoring—After the assessment occurs, the teacher records and interprets the results and adjusts teaching plans. There may also be sharing of findings with other teachers and delayed feedback to students.

4) Recording and Dissemination—Teachers report results formally, often for administrative/bureaucratic purposes.

Rea-Dickins points out that several decision-making processes and strategies are deployed at each stage. In addition, a teacher does not necessarily go through every stage; an informal observation, for instance, may only involve one or two stages.

A more “macro-level” model can be found in Hall et al. (1997). They studied teacher assessment at Key Stage 1 in an English LEA, and use a five-stage “developmental model” of teacher assessment to explain their interview data. According to this model, teachers planned assessments before the start of the school year, keeping in mind attainment targets set in the National Curriculum. They then observed students early in the school year (similar to what Airasian, 1997, calls “sizing up” assessment); this is followed by a third stage in which more specific tasks are set in light of the earlier observations and National Curriculum requirements. In the fourth stage, teachers continuously review their students over the school year; notably, this is more formal than earlier stages. The final leveling stage is summative; over the final 4-6 weeks, teachers assign a level of attainment to each student.

I would broadly describe each of these conceptions as “line” models of assessment; that is, they conceptualize assessment as a cognitive progression through several interactive steps. In this way, they seem relatively simple and straightforward. However, if we consider that, 1) as Rea-Dickins notes, each step involves other decisions and thought-processes, and that 2) these assessment models are in a sense nested (i.e., many small assessment processes occur in the context of larger ones, much as Woods, 1996, found with teaching decisions in general), then the real complexity of assessment thinking becomes more apparent. For example, at the planning stage, why does a teacher choose particular assessment activities? At the stage where the teacher interprets student “data,” against what kinds of criteria, norms, etc. are results judged? If a grade must be given, what considerations does the teacher take into account?

Below, I discuss three areas in which research has been done on such questions. The first two—implicit constructs and teacher assessment profiles—emphasize individual differences in assessment thinking, while the third—context influences—emphasizes the impact of external factors on the individual teacher’s thinking.

2.4.2 Implicit constructs

The importance of construct validity in testing ever since Messick’s seminal 1989 article has been extended in recent years to classroom assessment. Thus, researchers have become concerned about the implicit constructs that teachers use as they assess students; that is, “what do teachers look for when they are assessing?” and “what theory or ‘standards’ do teachers use when they make judgements and decisions?” (Leung, 2004). For example, in language teaching, a teacher who puts more weight on students’ accurate use of grammar than on their conversational fluency can be seen as using a construct of language ability based on a more structural rather than functional view of language.

There is a substantial literature (which I do not have space here to examine, but see, for example, McNamara, 1996) examining how raters for formal performance tests vary in what they consider to be salient aspects of a testee’s performance. Instead, I review here two studies that have looked at the constructs that teachers employ when assessing student performance.

Orrell (1995), studying Australian academics’ thinking as they assessed student writing assignments, found that they employed what was termed “reference models”—tacit, complex, and multifaceted mental constructs that were derived from discipline knowledge and experience. These constructs included 1) a concept of an ideal student performance on the task, 2) a concept of a range of possible student performances, and 3) knowledge of a grading scheme and its relationship with possible performances. Orrell draws similarities between these reference models and schema, adding that the models could be incrementally adapted with each new assessment experience.

Specifically related to language assessment, Leung & Teasdale (1997) examined the constructs that English primary teachers used as they assessed the language abilities of their English-as-Mother-Tongue (EMT) and EAL students. They found that the teachers shared similar understandings of general criteria when assessing speaking and listening; in particular, the construct of “native speaker-like” seemed to be widely held by the teachers, suggesting that the “native speaker” was seen as a norm. Also, Leung & Teasdale found that the constructs that teachers used were not all included in the National Curriculum level descriptors, suggesting a discrepancy between public assessment frameworks and what teachers are “supposed” to look for on one hand and teachers’ implicit constructs and what teachers actually look for on the other.

This last finding is of particular relevance to the increasing use of published assessment standards and frameworks (see Brindley, 1998, for a discussion of this phenomenon as it relates to language assessment); several studies (e.g., Davison, 2004; Arkoudis & O’Loughlin, 2004) have shown that because it is individual teachers who conduct assessment, they will have their own implicit constructs that will differ from published criteria and “community of practice” standards—one reason being simply that there is always individual interpretation of published standards occurring; teachers are thus always “mediators” of their own or external assessment criteria (Leung, 2004).

2.4.3 *Teacher assessment “profiles”*

Several studies have proposed that, where assessment is concerned, teachers can be categorized according to a typology of assessment “profiles” or orientations. The studies below propose different parameters or dimensions of such a typology, but they all suggest that teachers as individuals display clusters of characteristic assessment practices and thinking.

Edelenbos & Kubanek-German (2004), based on ethnographic data from a larger study, propose that three selected language teachers in their study can represent teacher assessment “types”. They discuss each teacher in three areas: attitudes towards language growth, overall attitude towards the larger project’s research

instruments (which involved assessment of students), and “diagnostic competence” (their term for the ability to interpret foreign language growth in individual children). The three types were:

Teacher 1—She was teacher-centered and emphasized immediate error correction. She was pro-active in carrying out the assessments required by the research, and also very active in her “diagnostic competence.”

Teacher 2—She was more “holistic”, seeing language growth as happening naturally through authentic interactions. She conducted all the required assessments but preferred the more creative ones, and she emphasized meaning-making when diagnosing student ability.

Teacher 3—He focused very much on correct grammar and pronunciation. He seemed more hesitant to do the assessment tasks, and he seemed to diagnose more generally rather than individually.

This study highlights how, as I have argued throughout this chapter, teachers access a wide range of thoughts in the process of assessing students; here, teachers’ beliefs about language and language learning (thus also relating to implicit constructs) are shown to play an important role.

Like Edelenbos & Kubanek-German’s “types,” McCallum et al. (1993) studied teachers under the National Curriculum assessment regime and found that they could be divided into 3 “styles”: Critical Intuitives, who disagreed with National Curriculum assessment methods and relied on their own informal, intuitive techniques; Evidence Gatherers, who gathered evidence of student achievement but avoided systematic assessment because they saw it as interfering too much with the teacher-children relationship; and Systematic Planners, who felt their teaching was enhanced by assessment and used it widely to help decision-making. Bachor & Anderson (1994) found similar results among primary teachers in British Columbia.

Samuelowicz & Bain (2002), in their study of academics in Australia, likewise propose a typology of orientations to assessment practice: theirs had six orientations that were differentiated in terms of the academics’ beliefs about such things as the nature of knowledge in their field, the role of assessment in teaching and learning, and the ways in which feedback from assessments should be used. Hence, for example, one orientation was labeled “reproducing bits of knowledge” (p. 184);

participants categorized under this orientation believed that knowledge in their field was atomized and needed to be recalled rather than transformed, that assessment was a means of making students study, and that assessment feedback should be used to alter teaching. Another orientation was labeled “transforming conceptions of the discipline/world”; academics with this orientation believed that knowledge taught to students had to be linked to students’ previous knowledge and had to be transformed by the student, that assessment was a means of guiding students’ learning, and that feedback should be used to challenge students’ understandings. These different orientations then led to different assessment practices.

Particularly relevant to this present study is Cumming (2001). Based on his interviews with highly experienced ESL/EFL writing teachers in different countries, he found that a teacher’s writing assessment practices were strongly related to whether the teacher took an “English for specific purposes” or an “English for general purposes” view of the writing curriculum. Teachers who saw their course content as aimed for specific purposes were much more focused on assessing specific competencies or behaviors, and thus used a relatively narrow range of assessment forms; in contrast, teachers who took an English for general purposes orientation were more likely to assess a wider range of achievements, in areas like not only language but also student self-confidence and acculturation into academic communities.

A few other studies have found variations in how teachers see themselves in the role of assessor. A consistent theme in these findings is that confidence in one’s judgments, or lack thereof, is a key difference between teachers. Reali et al. (2001), who investigated what Brazilian primary teachers considered as they assessed students’ written work, found four patterns in teachers’ reports to four texts they were given:

- 1) Some teachers looked mainly at surface level features and issues such as cohesion and style; in their reports, they rarely mentioned what they thought the students’ characteristics were.
- 2) Some other teachers also looked at formal aspects of the text but mentioned student characteristics more than the first group.

- 3) Teachers of the third pattern considered more subjective aspects of the text (e.g., “frankness”) and possible characteristics of the students, such as effort spent.
- 4) The last group often judged texts in reference to their own students and claimed to require more information about the student before being able to arrive at an evaluation.

Especially interesting is the finding that each group displayed progressively less confidence and security in their reports. The last group, particularly, wanted to know more about the student’s background and context before making a decision; this was attributed to a more constructivist view of teaching and learning.

Yung (2001; 2002), in his study of biology teachers in Hong Kong, found similar differences in confidence regarding judgments. He describes three teachers: the first had little confidence in his assessments of students and interpretation of public assessment standards, and thus felt it necessary to find faults with students so as to avoid being seen as too lenient by superiors; the second was much more confident in his assessments and interpretation of public assessment standards, and tried to use what he felt were student-oriented assessments; and the third had some practices similar to the second teacher but actually had little confidence and quite different motivations.

In her study comparing English language teachers in Hong Kong and Australia, Davison (2004) proposes a cline of teachers’ assessment orientations (see Appendix 1), which interestingly overlaps with findings from the previously mentioned studies. At one end is a technical orientation, whereby the teacher sees assessment merely as a matter of following published criteria. At the other end is an intuitive orientation, similar to those in McCallum et al.’s study, whereby the teacher makes judgments based on unarticulated references that are beyond analysis. In between are orientations that are somewhat of a mix between the extremes. Like the groups in Reali et al.’s study, teachers also varied in their focus on either text or student.

As these seven studies taken together suggest, teachers will vary in terms of their approaches to assessment (related to such things as their beliefs about teaching and learning, their professional experience, and their views of the subject matter), their confidence as assessors (although it should be noted that this is mainly in their

summative role), and their proportional use of private, student/context-focused intuitions versus public, performance-oriented criteria when making judgments.

2.4.4 Contextual influences upon teacher thinking in assessment

By all accounts of teacher thinking, including those related to assessment, contextual factors play an enormous role in shaping assessment decision-making. Below I discuss the impact of community and of government policy mandates, especially high-stakes testing, as these have been found to have significant influence.

Hall & Harding (2002) placed the schools in their study on a continuum between two endpoints. At one end were schools with an “assessment community,” in which teachers and administrators accepted and complied with National Curriculum assessment policies, were committed to group moderation, and shared a common language of assessment, among other characteristics. At the other end were schools with “assessment individuals,” in which teachers were reluctant to comply and resisted the mandated assessment policies, moderation was weak or did not occur, and there was uncertainty or confusion about assessment terms. Thus, a teacher’s immediate professional community can be considered to be one influence upon assessment thinking.

Community in a wider sense also seems to have an influence. Davison (2004) not only found the individual differences mentioned earlier, but also noted that teachers in each of the two cities (Melbourne and Hong Kong) reflected similarities in approaches and attitudes towards assessment. The Melbourne teachers all referred to the government-established standard criteria for assessment, while the Hong Kong teachers displayed much more diversity in their assessment processes and criteria. Davison attributes this to the presence or absence of mandated standards and consensus on criteria.

Perhaps the most discussed and studied contextual factor is government-mandated assessment policies. Several of the studies described earlier (e.g., Bachor & Anderson, 1994; McCallum et al., 1993; Rea-Dickins, 2001; Leung & Teasdale, 1997; Yung, 2001 & 2002) were actually investigations of teachers’ assessment

under new government policies about assessment, such as those allied to the English National Curriculum. It is clear from these studies that government policies on assessment, such as those requiring particular assessment practices or the use of assessment frameworks (see 2.4.2), can exert a strong influence upon teachers' classroom assessment, but ultimately those policies are mediated and adapted (or even rejected) by individual teachers.

One particularly significant policy is the installation of high-stakes testing at different points in the education system for not only the purpose of measuring student achievement but also the purposes of accountability and selection (Broadfoot, 1996). I will not examine the extensive literature on such testing's effects (also known as "washback" effects) upon teachers—see Alderson & Wall (1993), Bailey (1996b), and Cheng (2000) for reviews and discussion in language education—except to note that washback effects are complex; high-stakes tests often constrain teacher practices, yet teachers still hold on to their pedagogical beliefs and may work around such constraints (Wall & Alderson, 1996).

It is so far unclear, however, how government policies influence classroom assessment in EAP, as it is not centrally regulated like classroom assessment in the primary and secondary contexts, nor is it subject to mandatory high-stakes testing (a test like the IELTS⁶ arguably could have such an effect, but students on many EAP courses do not have to take the IELTS, and ones who do can opt for courses especially designed to prepare for the test); there is a fair degree of freedom as far as assessment in EAP is concerned, within the broad requirements set for universities more generally by government quality assurance policies for higher education.⁷

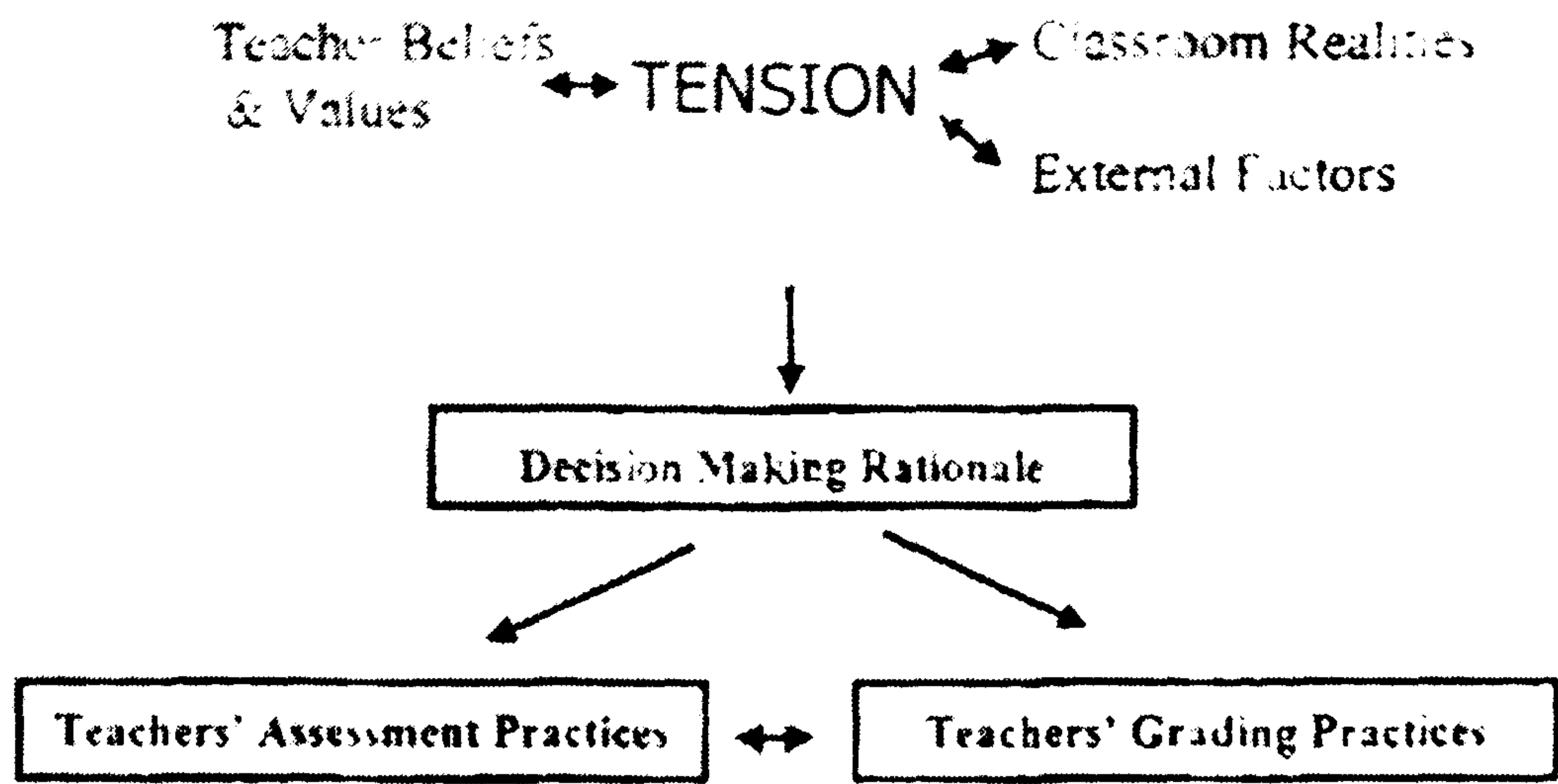
⁶ The IELTS (International English Language Testing System) is a high-stakes language test that students outside the UK are often required to take if they desire admission to a UK university or postgraduate course.

⁷ Broadfoot (1996) argues that governments are especially interested in assessment policy for primary and secondary education because they see it as a means of controlling the system and directing it towards government goals like economic productivity. I think the absence of explicit government policy regarding EAP assessment shows government's lesser stake in the area, although it is an open question as to how government policies for universities regarding quality assurance have an indirect influence.

2.4.5 A model of teachers' classroom assessment decision-making

Based on their study of primary and secondary teachers in Virginia, McMillan & Nash (2000) propose what I would describe as a “network” model of teacher decision-making in relation to assessment (Figure 2.3). I will give it some attention because elements of the issues discussed in the preceding sections appear in it.

Figure 2.3 Teachers' assessment decision-making (from McMillan & Nash, 2000, p. 11)



Analysis of interview data pointed to a variety of issues involved in assessment decision-making; these were grouped under five main themes: teacher knowledge, beliefs, expectations, and values; external factors (such as the existence of high-stakes tests); classroom realities (such as student absenteeism); decision-making rationale; and assessment practices.

As McMillan (2003, p. 36) notes:

Teacher internal beliefs and values were clearly cited as the most important influence on assessment decisions. What was particularly interesting was that these beliefs and values were not directly concerned with measurement principles. Even though the questions to the teachers focused on their classroom assessment practices, the most important reasons given were based on more general beliefs and values about teaching and learning. These more general beliefs were often voiced as a philosophy of education, and desired assessment and grading practices were consistent with this philosophy.

However, these were in tension with external factors and classroom realities, which forced teachers to use assessment practices that did not align with their beliefs and values. The researchers also found that teachers had difficulty giving reasons for using particular assessment practices: “[assessment decision making] was a highly

individualized, idiosyncratic process.” (McMillan & Nash, 2000, p. 38) Several reasons were postulated as to why this was:

- 1) educational measurement principles derived from standardized testing were not perceived as relevant to constructivist theories of learning or daily realities of teaching;
- 2) classroom assessment training was lacking;
- 3) external pressures like testing were pushing out alternative assessment approaches and forcing particular practices out of keeping with teachers’ desires;
- 4) classroom assessment decision-making is both an assessment decision and an instructional one, so pedagogical implications of assessments are heavily considered.
- 5) current ways of thinking about classroom assessment may be irrelevant to teachers’ pragmatic needs and pedagogical goals.

Finally, McMillan & Nash found that teachers used homework, quizzes, tests, performance assessments, participation, daily checks and informal observation. Teachers were also found to be creating and constantly revising their own assessments.

McMillan (2003, p. 38) concludes:

It appears that teachers are striving to reach a reasonable balance between their beliefs about education and learning on one hand and external factors on the other. This constant state of tension, exacerbated by high-stakes tests, characterizes teacher decision making about assessment and grading practices.

2.5 Summary of relevant insights and gaps in knowledge

The preceding sections have highlighted insights from both the general teacher thinking research and research specifically investigating teacher thinking in relation to assessment. Below, I summarize these insights but also discuss what I perceive to be the gaps that need to be investigated.

First, it is abundantly clear that beliefs and/or principles—not only about assessment but also about teaching and learning—have a significant impact on assessment practices; these beliefs and principles also play an important part in what can be characterized as categorizations or types of teacher approaches to assessment (see

2.3.1 and 2.4.3). However, while the general teacher thinking literature has shown how teachers call on beliefs and principles when working in the classroom, this has not included when assessing students in the classroom. On the other hand, most of the reviewed assessment studies that have examined assessment beliefs and principles have not done so in the classroom context; methodologically, they mostly used self-report interviews (see Table 2.1). The main problem with this is that the beliefs and principles thus found are in a sense abstracted for the interview context and not necessarily connected to classroom interaction; a discrepancy can therefore arise between espoused beliefs/principles and beliefs/principles in classroom practice (see Fang, 1996, for a discussion). It is unclear therefore as to whether or how such beliefs and principles are employed during *in situ* interactive assessment in the classroom.

Second, the literature makes clear that teachers engage in other cognitions besides calling upon beliefs/principles when assessing student performance. One key mental activity is that they compare student performance with their own implicit constructs or criteria (see especially 2.4.2 but also 2.4.1 and 2.4.3). The literature also suggests that teacher cognitions may be routinized (see 2.3.3). However, few studies have attempted to look at teachers' assessment cognitions more holistically; they have focused mainly on particular cognitions (like implicit constructs in Leung & Teasdale, 1997) or particular assessment practices (like grading written performance in Davison, 2004, and Reali et al., 2001), and they have often used somewhat experimental conditions, outside the classroom (like Orrell, 1995). Some studies have attempted to look, holistically, at assessment cognition in the classroom (Mavrommatis, 1997; Rea-Dickins, 2001; Edelenbos & Kubanek-German, 2004); these studies have provided valuable insights but are few in number and need to be bolstered.

Third, the general teacher thinking literature has pointed out important influences upon teacher thinking, such as contextual factors, training, and teaching experience (see 2.3.4). The assessment literature has also done so, at least with contextual factors like communities and government policies (see 2.4.4). Meanwhile, the assessment literature has put forth models (see 2.4.1) of the assessment process and what teachers think about at each step in that process (or, more appropriately,

processes). However, these two general findings have not been integrated—that is, there is so far no model of how those important influences as a whole impact teachers’ assessment thinking—except for the model proposed by Macmillan & Nash (2000). That model begins to adumbrate the influences and how they affect teacher thinking, but as the authors admit, it is exploratory and much more work needs to be done in this area.

Finally, one glaring gap in current knowledge, applicable to all three points above, is the lack of studies in contexts outside primary and secondary schooling. Of the three studies in this review that did not involve primary and secondary education, two (Orrell, 1995; Samuelowicz & Bain, 2002) involved the academic context but not EAP, while one (Cumming, 2001) examined writing assessment only. There is thus much room for empirical investigation in a variety of contexts, including EAP.

2.6 Summary

In this chapter, I have located my study in the classroom assessment literature and explicated the definitions of classroom assessment and practices that I use. I have also given an overview of key concepts from the general literature on teacher thinking and reviewed previous research into teacher thinking in assessment. Finally, I have summarized the main insights gleaned from this review of the literature and pointed out gaps in current knowledge.

In the next chapter, I begin to set out the methodology I used to conduct an investigation that sought to fill some of those gaps.

Chapter 3 Methodology: Overview & Stage 1

3.1 Introduction

In this chapter, I begin to present the methodology underlying the empirical basis of this dissertation. In doing so, I move from a general discussion of my philosophical position to an explanation of the principles guiding the overall design of this study. I then focus on the rationale behind the decisions involved in the planning and conduct of the Stage 1 data collection. This is followed by an explanation of how the Stage 1 data was analyzed. The last section of this chapter deals with ethical issues arising in Stage 1. The rationale underlying Stage 2, the analysis of its data, and ethical issues arising from it will be dealt with in chapter 5.

I must mention at this point that while for the purposes of presentation I have tried to describe my methodology in a generally linear fashion, the reality of the process I went through in planning and executing the study reflected what Bryman & Burgess (1994, p. 2-3) state about the research process:

Indeed, research seldom involves the use of a straightforward set of procedures. Instead, the researcher has to move backwards and forwards between different sequences in the research process. For example, in designing a project, consideration needs to be given to the end-point and the concepts and theories that will be used in data analysis. Similarly, in terms of data collection, reference has to be made to the comparisons and contrasts that may be uncovered during a project. On this basis, there is not a sharp divide between different aspects of the research process in practice.

3.2 Philosophical position

It is good medicine, we think, for researchers to make their preferences clear. To know how a researcher construes the shape of the social world and aims to give us a credible account of it is to know our conversational partner. (Miles & Huberman, 1994, p. 4)

Crotty (1998) describes a philosophical position as consisting of views about reality, knowledge, and methods (corresponding to ontology, epistemology, and methodology, respectively). While the common division of such positions by social science methodology texts into broad categories like positivism/post-positivism, interpretivism/constructionism, and critical theory has heuristic value, I agree with Miles & Huberman (1994, p. 4-5) that the boundaries between these categories have become blurred, and that "... in the actual practice of empirical research. ... all of

us—realists, interpretivists, critical theorists—are closer to the center, with multiple overlaps.” (p. 4-5)

While I hesitate to claim strong adherence to a particular philosophical system, I have found that my personal views probably most closely align with what has been called critical realism (Porter, 2002) or transcendental realism (Miles & Huberman, 1994).¹ It is ontologically realist, particular in regards to social reality; given that human understandings and actions display a great deal of regularity and pattern rather than randomness, it follows that structures beyond the individual must exist in some way (Porter, 2002) “Unlike researchers in physics, we must contend with institutions, structures, practices, and conventions that people reproduce and transform. Human meanings and intentions are worked out within the frameworks of these social structures—structures that are invisible but nonetheless real.” (Miles & Huberman, 1994, p. 4) In terms of epistemology, critical realism posits that those structures and how they influence and are influenced by human agency can be empirically examined. However, it also sees reality as an open system; thus, what may be discovered and explored are tendencies in social life that have a contingent nature, rather than unchanging cause-effect relationships. Critical realism has been labeled as post-positivist (Guba, 1990)—although the label itself oversimplifies a broad, complex, and amorphous approach (Phillips, 1990) that encompasses a variety of philosophical positions; Porter and Miles & Huberman portray critical realism as being situated between positivist and hermeneutic social science traditions; it attempts, on one hand, to affirm the importance of subjectivity and meaning-making, while on the other hand to somehow account for observable regularities in social life.

¹ My “skepticism” towards various *philosophical* systems of thought is rooted in my *theological* system of thought, which is derived from the Reformed tradition of Christianity. A key Reformed tenet, total depravity, refers to the belief that all human faculties are ultimately influenced by human sin and are thus fallible and incapable of absolute knowledge. Francis Schaeffer, a 20th-century Reformed thinker, has argued that philosophy unattached to theology always leads to a search for meaning and value in irrational sources; thus, for example, various thinkers have sought ultimate meaning in music, poetry, drugs, and so forth (Schaeffer, 1968). I bring this up because I think Roy Bhaskar, who is credited as the main thinker behind critical realism, can be seen as doing this; he has, after a career of setting forth a philosophical basis for the natural and social sciences, argued that ultimate meaning and value can be found in what can be termed a kind of New Age spirituality that involves concepts like karma and reincarnation (Bhaskar, 2000; Hartwig, 2001). Hence, while I find that critical realism has elements that I can affirm from my theological standpoint, I still hold that paradigm position rather lightly, only as a heuristic tool to provide a framework for research and not as a source for ultimate values.

This critical realist position underlies several assumptions I made in approaching the investigation of teacher thinking and assessment. First, I assumed that there were patterns of teacher action that could be described as assessment practices; while the term is a construct, the construct attempts to describe an underlying reality in the classroom (similarly to how the construct of language proficiency attempts to describe a real aspect of an individual). I also assumed that teachers had thoughts guiding those practices, and that those thoughts could be expressed, discussed, and examined. In addition, I assumed that such thinking did not occur in an individual vacuum, so to speak, but occurred within and interacted with a socially real context. This meant putting a high priority on understanding the context(s) in which teachers conducted assessment. This emphasis on context went hand-in-hand with a belief that educational institutions and systems are configured in such a way at this period of time that a close examination of phenomena in context actually has wider generalizability and relevance outside that local situation.

While the search for patterns and regularities may suggest the use of quantitative methods under experimental conditions, critical realism's concern with social life actually means that a great deal of research taking this stance uses qualitative methods in more naturalistic settings (Guba, 1990; Miles & Huberman, 1994; Porter, 2002).^{2,3} The present study follows in this approach. While a critical realist position allows and perhaps lends itself to the mixing of qualitative methods with quantitative methods like surveys or systematic classroom observation (Miles & Huberman, 1994, p. 40-43), I chose to deploy only qualitative methods in my study; the rationale for this and other design decisions are the subject of the next section.

3.3 Principles guiding the overall research design

Table 3.1 gives an overview of the complete study's overall design. In this section, I explicate the principles that guided the development of that overall scheme.

² Guba is actually a critic of critical realism, but I think he describes it fairly here.

³ I also think Yin (2003), while not explicitly tied to critical realism, shares many of the same assumptions.

Table 3.1 Overview of research design

Stage 1	
Research questions: RQ1. What are teachers' definitions of assessment? RQ2. What are the classroom assessment practices of English language teachers in the EAP context? RQ3. What cognitions underlie these practices? RQ4. What are the sources of these cognitions (e.g., initial teacher training)? Strategy: Case study Data collection methods: * Classroom observation (RQ2) * Interviewing (RQ1) * Stimulated recall (RQ3, RQ4) Sample: 2 teachers teaching an insessional EAP course for 1 term in a university language center Timeline: Data collection Jan 2004-April 2004 Analysis (inc. in field & writing) Feb 2004-Dec 2004	
Stage 2	
Research questions: RQ5. How do English language teachers think they can increase the quality of their impressionistic knowledge of a student's language abilities? RQ6. How do teachers who have experienced managerial changes like a more explicit syllabus and more explicit summative assessments feel such changes have affected their assessment practices? RQ7. Do teachers use stereotypes and projection when they conduct interactive assessment of students? Strategy: Interview Data collection method: Focus group interviewing (RQ5, RQ6, RQ7) Pilot: 3 doctoral students in TESOL with EAP experience in EFL contexts Sample: 3 focus groups, each consisting of teachers from a university EAP center or program Group 1: 4 participants; Group 2: 3 participants; Group 3: 5 participants Timeline: Data collection March 2005-June 2005 Analysis (inc. in field & writing) March 2005-June 2005	

As Table 3.1 shows, the study consisted of two stages, each with a distinct but related set of research questions. This was based on the idea of progressive focusing (Stake, 1995, attributes the concept to Parlett and Hamilton, 1976.)—starting with general research questions and then allowing the data collection and analysis to bring further, more specific research questions into focus, which are then studied in turn. Marshall & Rossman (1995, p. 43) describe a similar principle of maintaining flexibility in qualitative research, so that experiences in the field can begin to “clarify the relevant themes and patterns.”

I thought this principle was particularly important for my study because, on one hand, relatively little research had been done in the intersection of teacher thinking and assessment, so a fairly wide exploration of the territory needed to be done first; hence, the Stage 1 research questions (RQ1-4) were of a fairly broad and descriptive nature. On the other hand, such an exploration alone risked superficiality, so a more focused examination of key issues was needed to develop deeper insights that would contribute more substantially to what is currently known about teacher thinking in classroom assessment. Hence, the Stage 2 research questions (RQ5-7) were more precisely stated and involved specific issues raised by the Stage 1 findings.

As Table 3.1 also shows, each stage involved a different strategy and set of data collection methods. By strategies I mean general plans about how to go about answering questions, and by methods I mean specific ways of gathering data; a strategy is therefore effected through a method or methods. Thus, for example, case study would be a strategy while non-participant observation of a case would be a method. The decisions about strategies and methods were partly guided by the principle of fitness for purpose; strategies and methods should fit the research questions. Zelditch, in Marshall & Rossman (1995), gives two important criteria for judging the appropriacy or fitness of a strategy. The first is informational adequacy, whether the strategy will generate enough quality data to allow the researcher to answer the questions. The second is efficiency, whether the strategy will provide sufficient data for the least cost to researcher and participants. These two considerations weighed heavily in my planning of the two stages of my study.

In order to follow the progression of the study as it unfolded, I have divided my discussion of each stage into separate chapters. Stage 2 is dealt with in chapter 5. Below, I elaborate the rationale underlying the decisions made in the Stage 1 data collection.

3.4 Rationale for design and conduct of Stage 1 data collection

3.4.1 Stage 1 research questions

As I have mentioned, because previous research in the intersection of teacher cognition and classroom assessment still seemed very exploratory, I decided to begin my study with some general questions:

RQ1. *What are teachers' definitions of assessment?*

RQ2. *What are the classroom assessment practices of English language teachers in the EAP context?*

RQ3. *What cognitions underlie these practices?*

RQ4. *What are the sources of these cognitions?*

RQ3 actually provided the main impetus for this study, while the other questions fulfilled important functions: RQ1 served as a check on the match between researcher and teacher views (see 4.3), RQ2 operationalized teacher thinking in assessment (i.e., served as a bridge between observable phenomena and teacher cognitions), and RQ4 sought to explore the background to teacher cognitions.

3.4.2 Case study strategy

I chose a case study strategy to answer these questions primarily because I wanted to look at assessment practices and related thinking in the natural context of the classroom, and the suitability of case study for researching phenomena in context is widely recognized (Bogdan & Biklen, 1998; Marshall & Rossman, 1995; Stake, 1995; Yin, 2003). While an experimental or survey strategy would have given me some purchase on these questions, I was aware from the teacher thinking literature that contextual factors play a vital part in teacher cognition (see 2.4.4). In addition, I saw classroom assessment as a kind of situated cognition (Salomon, 1993), and so it was important to include the situation in any research account. In terms of Zelditch's criteria, data of the nature I wanted (contextualized) would be most sufficiently and efficiently provided by a case study strategy. In choosing this strategy, I also was in line with teacher thinking research more generally; because the strategy attempts to account for context and allows in-depth exploration, it aligns well with a view of

teachers as thoughtful professionals (see 2.3) and thus has an established place in teacher thinking research methodology.

One important point that the case study methodology literature emphasizes is that the case needs to be delineated clearly. In Stage 1, I defined a case as a teacher teaching a course for one term. Teacher thinking research has often taken the teacher alone as a case, looking at his or her thinking in a holistic manner to include multiple courses, different groups of students, more than one academic term, etc. (Elbaz, 1983; Woods, 1996). However, I bounded the case further because classroom assessment by definition occurs for a teacher in a classroom with students who are gathered for a course, and a course has certain structural constraints—duration of each class session, duration of a term, syllabus, number of students, etc.—that the teacher must work within. In other words, assessment occurs in a specific context, and my definition of the case reflected this. Bounding the case in this way also allowed me to compare cases (Miles & Huberman, 1994; Yin, 2003) as an analytic tactic; given ostensibly the same course (i.e., a fairly similar set of contextual factors), how were teachers similar or different in their assessment practices and thinking? The class in these cases was an EAP insessional course, which covered English listening, speaking, reading, and writing for academic purposes and which occurred during the fall, spring, or summer term.

3.4.3 *Data collection methods*

Within a case study framework, I employed a few specific data collection methods. Below, I discuss their rationale and how these were implemented in practice. In doing so, I occasionally refer to the two EAP teachers involved;⁴ I elaborate on sampling in the succeeding section.

I. Non-participant observation

This is a mainstay tactic for data collection in case study research (Stake, 1995; Yin, 2003), and has often provided the basic foundation for further inquiry into teacher thinking (Borg, 1998; Elbaz, 1983)—teachers are asked to explain their observed behavior in the classroom. Part of the importance of observation in my study is

⁴ Throughout the following chapters, the two Stage 1 teachers are referred to as Teachers 1 and 2, or CTA and CTB; in transcript excerpts and references they are usually referred to as A and B.

connected to my definition of assessment; since I took a broader view of assessment to include assessment during pedagogical activity, observation of pedagogical activity was required.⁵

I observed the teachers' classes by arriving a few minutes before class (usually), setting up my videocamera⁶, sitting in a corner of the classroom, and taking field notes—what Stake (1995) calls “qualitative observation”, as opposed to quantitative observation using a schedule of predetermined categories (see Appendix 2 for an example of field notes I took); during the class time, I also collected from either the teacher or the students near me any handouts that were issued. The videotaping was done for two reasons: 1) it provided a record that I could view again to confirm what I had written in my notes, and 2) segments of the recording served as the stimulus for the recall sessions (see III below).

Observer effects were apparent in two senses. First, I brought my concern with assessment practices to the process of observation—“there is more to seeing than meets the eyeball” (Phillips, 1990, p. 34)—and so my notes focused on this aspect rather than others, such as classroom discourse per se or student-student interactions. Second, my presence did influence classroom events. Sometimes this was obvious; for example, the teachers would occasionally come over to me and briefly talk with me during the class, usually to comment about the activity students were doing. In addition, the teachers told me explicitly that my study sometimes weighed on their thinking; one teacher said he had been thinking a lot about feedback recently, perhaps because of my research, and the other teacher said he felt he was being more prepared for class because he knew I was observing. Other times, the influence of my presence was less obvious and could only be guessed at; for example, in one of the teachers' classes, I suspected that some of the students intentionally picked seats so as to have their backs to the videocamera (see ethical issues in 3.6).

⁵ It is significant that research taking a narrower view of assessment practices often has relied on non-observational techniques, such as survey or interview (Cheng, Rogers, & Hu, 2004; McMillan, 2003), while research taking a broader view usually invests in classroom observation (Edelenbos & Kubanek-German, 2004; Rea-Dickins, 2001). I think this is because a narrower view of assessment as homework, quizzes, and tests, being less frequent or occurring outside class, does not lend itself to efficient investigation through extended classroom observation.

⁶ The first classes for each teacher were not recorded, as the videocamera was as yet unavailable to me from my department

Within post-positivist paradigms, objectivity is a “regulatory ideal” (Phillips, 1990)—a goal that guides research practice and towards which it aims—and so the observer effects mentioned above need to be accounted for (although not removed in the name of a nonexistent absolute objectivity). Regarding my bringing particular concerns to the act of observing the classroom, it is recognized in postpositivist paradigms that “while we must be aware of the role played by our preconceptions in influencing our observations, and while we have to abandon the view that observation is ‘neutral’ or theory free, [this does not mean that] we cannot decide between rival claims and cannot arrive at consensus about which viewpoint (or which observations) seem to be most trustworthy under the prevailing circumstances.” (Phillips, 1990, p. 35) I think this is helped by keeping observations low-inference; as the example field notes in Appendix 2 show, I wrote down classroom action on one side of the paper—e.g., “Teacher 1 asks SA to give answer to homework” or “Teacher 2 turns towards the SB-SC-SD group and observes them”—and kept higher-inference interpretive memos on the other side—e.g., “Does Teacher 1 see SA’s silence as ignorance?” or “I should ask Teacher 2 about his observing groups from a distance.” It is these low-inference observations that build towards my higher-inference interpretations; as Punch (1998) points out, this is a fundamental process in both qualitative and quantitative research (see also conduct of analysis in 3.5).

Regarding the second observer effect, I think my ongoing presence in the classroom—I attended 16 of Teacher 1’s 20 classes and 15 of Teacher 2’s 20 classes—helped offset some possible problems. For example, my stay in the classroom allowed me to be somewhat more accepted by the students; while some seemed particularly aloof with me at the beginning, they seemed to become friendlier as the term went on. The main concern seemed to be with the videotape (see ethical issues in 3.6), so I tried to make it clear during the class that I was keeping the camera focused on the teacher. Also, while there were instances where the teachers talked to me during class, I think I was present long enough to get a good idea of the teachers’ typical classroom practices, including those related to assessment. In addition, spending an extended period of time in the classes and seeing the teachers regularly allowed me to build rapport with them.

It could be argued that, based on the teachers' admittal that my presence had an influence on their thinking and preparation, what I observed was not typical of their "regular" practice. In response, I would argue that a) it would have been difficult to sustain innovatory practice for the whole term, considering that they had other classes to teach and other responsibilities to fulfill; besides, from the data, it was evident that some practices were innovative to a degree (for example, using a new test for diagnostic purposes), but most practices were "tried-and-true"; b) if the teachers did think more about the feedback they gave or prepared more, these seem to be positive effects, and it seems more important to observe the teachers at their best, for the sake of ethicality, rather than to hope to catch them at their worst, in the name of "objectivity."

II. One-to-one interviewing

Interviewing can refer to both an overall research strategy and to a plethora of specific data collection techniques (Bogdan & Biklen, 1998; Marshall & Rossman, 1995; Punch, 1998). For consistency, I use "interviewing" to refer to the strategy and specific kinds of interviewing (e.g., one-to-one, focus group) to refer to the methods. I take up general issues and debates related to interviewing in my discussion of the Stage 2 strategy in chapter 5; here, I only wish to mention that in Stage 1, one-to-one interviewing—like observation, another main source of data when employing a case study strategy—was used to gather information about: the teachers' professional backgrounds, aspects of the language center that they worked in, background on the course they were teaching, and the teachers' general views about assessment.

The interviews were conducted after the initial class observation with each teacher; these were tape-recorded and later transcribed for analysis. For the sake of accuracy and fair representation, I gave copies of these (and the stimulated recall transcripts) to the teachers for them to check; they endorsed all the transcripts, with only occasional minor changes.

III. Stimulated recall

After nearly every (with one teacher) or every other (with another teacher) class session, I sat down with the teacher in the same or a nearby classroom and, using the

videotape player and television that the room was equipped with, replayed videotape segments of classroom action that I had decided during the class observation may have involved the teacher's assessment of a student's language ability (see 3.4.4 for my selection criteria). As the tape played, I asked the teacher to try and recount what had been going through his mind at that time; either of us could stop the tape at any time and discuss further any issues raised. If an assessment practice was involved, I also asked about the original source of the practice. An example of the taped classroom action alongside the teacher reporting his thoughts is provided in Appendix 3.

Because this data elicitation method—usually termed stimulated recall or retrospective report—is crucial to the design of my study, I shall discuss it at some length; in particular, I wish to respond to the main criticisms leveled at the method.

1. Teacher thinking research and stimulated recall

Besides stimulated recall, Clark and Peterson (1986) give four other common methods that have been used to study teacher thinking:

- 1) thinking aloud—the teacher verbalizes his or her thoughts while (as opposed to after, as in stimulated recall) doing a task like planning a lesson (for examples, see Davison, 2004, and Orrell, 1995).
- 2) policy capturing—the teacher is presented with descriptions (of, for example, students, hypothetical teaching situations, and the like) and asked to make decisions or judgments about each description, often on a Likert scale. The data is then used to produce mathematical models of the teacher's "policy" regarding those kinds of judgments.
- 3) journal keeping—the teacher keeps a written record about the topic of research interest (e.g., planning).
- 4) repertory grid technique—the teacher is presented with cards containing single words or statements about the area of research interest and asked to indicate which are alike or different and why. The resulting groupings and rationale ("constructs") are then put into a grid format and analyzed to show the relationships between constructs (for an example, see Leung & Teasdale, 1997).

I would add to this list 5) questionnaires and 6) one-to-one or other types of interviews; many studies (see examples in Table 2.1) have used a combination of all of these.

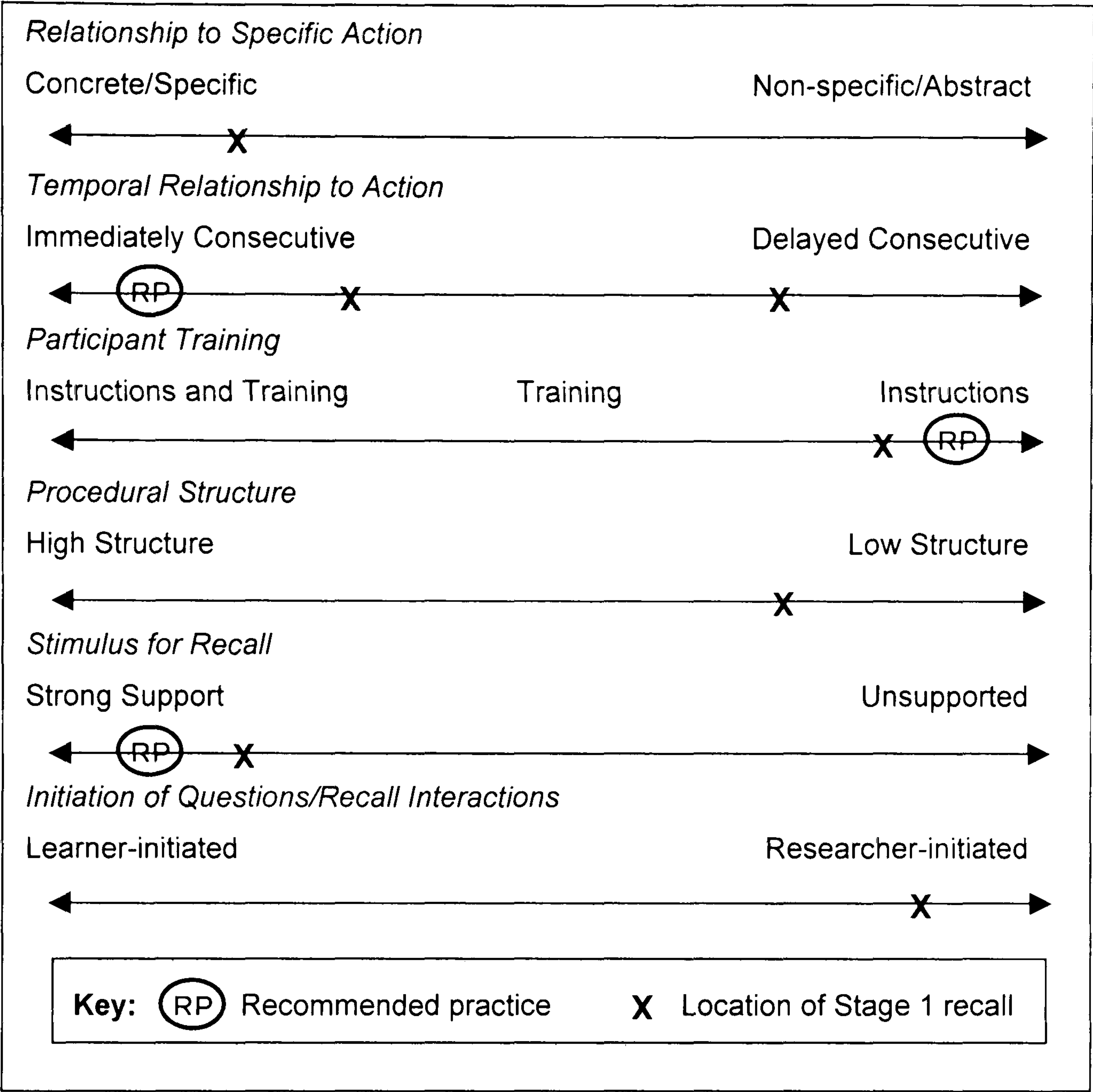
I chose stimulated recall mainly because it fit well with my case study strategy. Such a strategy precluded methods that required more experimental designs and/or larger sample sizes, such as thinking aloud, policy capturing, repertory grid technique, and questionnaires. As for the other techniques, I thought journal keeping and interviewing (by themselves) would elicit data at too high a level of abstraction/generalizability, without accounting for the situatedness of classroom assessment; the main appeal of stimulated recall vis-à-vis these two other methods is that it uses a record of actual classroom events to help the teacher remember his or her thought processes in specific situations.

Stimulated recall may seem incongruous with case study; after all, the procedures involved are somewhat “unnatural” insofar as they are typically used only in research and not professional practice. Indeed, much of the research using retrospective reports, particularly in cognitive psychology and second language research, has used experimental designs (Ericsson & Simon, 1993; Gass & Mackey, 2000). However, as Calderhead pointed out as early as 1981, a significant number of teacher thinking studies have employed stimulated recall within classroom-based research to elicit data on actual teacher thought processes. The technique may seem more congruent with case study if it is seen as a kind of interviewing except with the added advantage of the stimulus as a memory aid.

2. Parameters of stimulated recall

As with many methods, stimulated recall designs can vary widely. Gass & Mackey (2000), building on a classification scheme for introspective research developed in Faerch & Kasper, lists several parameters along which stimulated recall arrangements can be varied, producing different designs (Figure 3.1).

Figure 3.1 Classification scheme for stimulated recall parameters (from Gass & Mackey, 2000, p. 49, after Faerch & Kasper)



In Figure 3.1, I have also marked approximately on these parameters where Gass & Mackey recommend designs should fall (when mentioned) and where I believe my design falls. I discuss each parameter below.

a. Relationship to specific action

This refers to the specificity or concreteness of the action about which the participant is asked to recall. For example, asking about cognitions when writing a particular essay in a second language would be on the specific end, while asking about language learning strategy use in general would be abstract. I believe asking a teacher to recall thoughts during assessment practices like observing students doing classwork or giving presentations falls on the concrete end, although their occurring within the teacher’s ongoing “stream of consciousness” during classtime may make it less concrete or clearly delineated than in an experimental task.

b. Temporal relationship to action

This refers to when—immediately or some time afterwards—the recall session occurs in relation to the action being recalled. Gass & Mackey, echoing Ericsson & Simon (1993), recommend that the recall session be carried out as soon as possible after the action, primarily because of concerns about memory decay with long delays and the resulting likelihood that something besides the original thought processes, such as inferences or post hoc rationalizations, will be reported instead (see discussion below about criticisms). In Stage 1, since immediate recall was impossible under normal class conditions, I attempted to schedule the recall sessions as soon after the classes as possible. One teacher was available after nearly every class; however, the other teacher was only available after the second class session of every week, leaving a two-day delay between action and recall; I felt this was unavoidable, given the teachers' busy schedules. Overall, I thought these arrangements were nearly ideal and very generous on the teachers' part.

c. Participant training

This refers to whether participants receive instructions only or also receive training in doing a stimulated recall. According to Gass & Mackey, although there is not yet any conclusive empirical evidence either way, they recommend providing as little training as possible to avoid training effects or increased researcher input into the data. With my two case study teachers, I only provided instructions; they seemed quite able to provide recall information without training, although this may have led to the low number of instances in which they took the initiative to stop the videotape to talk (see below about initiation of interactions).

d. Procedural structure

This refers to how the recall data is collected; for example, having a participant choose answers on a multiple-choice questionnaire would reflect high structure, while an open-ended interview in which the participant chooses what, when, and how much to report would reflect low structure. I located my Stage 1 recall somewhat towards the low structure end: my usual question to the teachers when showing a video segment to them was “what was going through your mind at that time?”, so this was relatively open-ended but within a researcher-selected situation.

e. Stimulus for recall

This refers to the nature of the stimulus provided to the participant. Support can range from no stimulus at all to multiple forms of stimuli, such as video or audio tape, computer-captured data, written documents, and/or other artifacts. Gass & Mackey recommend using as strong, and as many sources of, support as possible, such as videotape with transcript. I used videotape as support for the teachers in Stage 1. This can be considered to be relatively high support, although as Gass & Mackey point out, the seeming strength of support cannot be assumed, since people have differing responses to similar stimuli; Fuller & Manning (1973), cited in Calderhead (1981), add that teachers viewing videotapes of their lessons are seeing the lesson from a different perspective/angle and initially tend to be distracted by their own appearance (something I found to be true in my study, also). These caveats aside, videotape seemed to provide the strongest and most efficient form of support given the circumstances of my study.

f. Initiation of questions/recall interactions

The last parameter refers to who chooses and interacts with the stimulus episodes; it can be the participant, the researcher, or both. Gass & Mackey point out that there are strengths and weaknesses with each option. Participant choice/interaction can prevent the researcher from unduly influencing the recall report but may lead to unfocused reporting or low amount of data (Gass & Mackey give an example from their own research, in which both participants and researchers could choose the episodes to discuss; in the end, only 10 percent of the replays in their study were participant-initiated.). Researcher choice/interaction can focus the report data but may “lead” the participant into responding a particular way (Calderhead, 1981) or put the participant in the position of having to report on something he or she might not recall. Gass & Mackey conclude that the optimal choice depends on the research questions. In my study, in accord with my paradigmatic position, my specific concern with classroom assessment practices, and my research questions, I was the one who chose the episodes; I also told the participants that they could stop the videotape during the segment whenever they wanted, but they rarely did so, so it was usually I who stopped the tape to allow time for the teacher to elaborate and for

myself to ask follow-up questions, particularly about the sources of their practices and thinking.

3. Criticisms of stimulated recall and response

Stimulated recall relies on the participant's memory and introspection to give the researcher access to essentially unobservable internal thought processes; this dependence on self-report has opened the validity of stimulated recall and other such introspective methods to substantial criticism. In this section, I wish to respond to what I judge to be the main arguments against the method; in so doing, I draw heavily on Ericsson & Simon's (1993) book-length examination of verbal reporting.

Before dealing with those arguments directly, I want to mention two points in support of self-report generally. The first is that, as has been pointed out in the methodological literature, self-report actually is the foundation of many methods, even those used in so-called "harder" scientific research like brain scanning, which depends on participant self-report to make links between mental processes and physical parts of the brain; thus, categorically ruling out introspection forecloses on a very wide range of methods, not only stimulated recall. Second, it seems to me that the more self-reports are depicted as untrustworthy, the more privileged the researcher's position becomes; in other words, when a researcher says that a participant was not really thinking what he says he was thinking—assuming honest intention to report on the participant's part—this implies that the researcher is able to see beyond the participant's "false consciousness" to a greater truth. It is particularly true in this situation because it is the participant's own thoughts that are being called into question. This line of thinking strikes me as rather untenable in light of 1) ethical arguments for greater equality and fairness between researcher and participant and greater reflexivity on the part of researchers, 2) efforts to decrease the research-practice gap by bringing the researcher and practitioner communities closer together (Bickel & Hattrup, 1995; Ellis, 1998; Huberman, 1990) (it seems rather difficult to build a relationship when one party takes a more privileged position), and 3) deflated epistemological claims for science (Phillips, 1990). This is not to say that self-reports should be naively taken at face value, but only to say that researchers should take enough care to avoid making a priori assumptions about the untrustworthiness of self-report.

A survey of the methodological literature on stimulated recall provides several criticisms of the method. Some are minor and can be dealt with straightforwardly. For example, it has been argued that teachers will feel anxious or distracted by their appearance when viewing a videotape stimulus; this can be overcome by building rapport between teacher and researcher and by letting teachers become familiar with the procedure (Calderhead, 1981). However, other criticisms of the method are more substantial; below, I outline the main arguments and give a response.

One common argument is that stimulated recall requires memory, and memory can fade quickly. As I mentioned earlier under temporal relationship to stimulus, recommended practice is to conduct the recall as soon after the original event as possible. The actual time limit before significant loss of recall quality has not been settled; Bloom found recall to be highly accurate within 48 hours of the event, while Cohen and others are cited as having found that most memory loss may occur shortly after the event, so that delays of from three hours to three days may have similar results (both cited in Gass & Mackey, 2000). This presents a validity concern for me regarding data from the teacher who was available to do a recall only after the second class session of each week, as I mentioned earlier; keeping in mind the possibly greater inaccuracy of recalls about segments from the first class session of the week, I took two measures with this teacher: 1) in data collection, I tended to have the teacher recall segments from the second session more often than the first;⁷ 2) in data analysis, I checked to make sure that if any of the supporting evidence for a finding came from recalls of the first session of the week, that there was also supporting evidence from recalls of the second session of the week. This follows the advice of Miles & Huberman (1994), who recommend strengthening validity of analysis by taking the quality of data into account.

The second and I think more substantial criticism of stimulated recall is that the thought processes that the researcher wishes to examine may be so tacit or automatized—having perhaps been developed through experience or trial and error (Calderhead, 1981)—as to be inaccessible to the participant and unable to be recalled.

⁷ This may have had the effect of emphasizing assessment of speaking and listening, since that was usually the subject of the second session. I add this caveat to my findings in chapter 4.

In a similar vein, Nisbett & Wilson (1977) argue that higher-order cognitive processes cannot be accessed, and that what recall participants are reporting are not actual thinking processes but are causal theories that the participants regard as rational explanations for the products of their thinking; in other words, the reports are after-the-fact rationalizations of their behavior.

Several points can be made against this criticism. The first is that I think teacher behaviors and their related thought processes—including those linked to classroom assessment—vary in degree of automaticity. Ericsson & Simon (1993) generally defend recalls but assert on theoretical and empirical grounds that thought processes underlying automatic behaviors cannot be recalled, simply because they never went through short-term memory in the first place (i.e., they bypassed memory and thus cannot be remembered). An easily recognized example of such automatic thinking is “highway hypnosis” (Natsoulas, 1970, cited in Ericsson & Simon, 1993), in which drivers cannot remember anything about the last 20-30 miles of highway driven because they have been thinking about other subjects while performing the task of driving automatically. While I recognize that some teacher behaviors are likewise highly automatized, I would hesitate to compare most teachers’ classroom behavior and thinking with this level of automaticity (although the oft-heard-in-staff-lounges idea of teaching “on autopilot” when tired or ill may belie this!); rather, I think there is a range of automaticity, with constantly repeated minor actions (e.g., passing out handouts) on one end and relatively rarer substantial actions (e.g., explaining at length a language point raised by a student) at the other end, with other actions (e.g., grouping students) being sometimes automatic and sometimes deliberated. I think this is likely true of classroom assessment, also. For example, nonverbal cues such as furrowed brows as showing a student’s lack of understanding may involve a nearly automatic recognition process, which Ericsson & Simon argue cannot be recalled, while assessing a student’s in-class presentation may be less automatic and thus able to be recalled. All this is to say that I expect the use of stimulated recall will enable me to capture a good deal but not all of the thought processes involved in classroom assessment.

Having hedged my claims, I would also assert that some tacit, unconscious, or automatic—I use the terms somewhat interchangeably, since they all relate to

thinking that is not consciously noticed—processes may actually be subject to at least some degree of recall, for two reasons. The first is that those processes may be more conscious—and thus more likely to leave a memory trace—than they appear. In their critique of introspective methods, Nisbett & Wilson (1977, p. 240) give the example of creative workers’ (e.g., artists, mathematicians, etc.) sudden flashes of insight; that is, many creative workers often talk about encountering and considering an intractable problem and then not thinking about it for a period of time; then the solution suddenly “pops” into mind. There is no report of any conscious thought behind the sudden insight. However, Ericsson & Simon (1993) revisit this topic and, citing empirical studies that more closely observed problem-solving and creative thought processes, argue that in fact there is still a good deal of gradual and conscious thought involved; what happens is that intermediate steps, often in which possible solutions are examined and discarded, may be unplanned (e.g., thinking while driving) and of short duration and are hence easily forgotten.⁸ What I am arguing here is that there are likely truly automatic/tacit/unconscious thought processes that cannot be recalled by the teacher, and there are thought processes that may seem automatic/tacit/unconscious to the teacher but that can actually be recalled, at least in part.

The second reason I think some tacit/unconscious/automatic processes may be recalled is because of what Ericsson & Simon call their “regeneration hypothesis”:

In many cases where the same behavior is required again and again, the conscious level of control is changed to a monitoring role, and the corresponding memory trace may become weak or disappear. ... Even in situations with rather full retrospective reports, subjects often have difficulty in retrieving the corresponding episodic memory. However, this doesn’t imply that they cannot give valid reports on “what they must have done.” When repeating the same task, like multiplying two 2-digit numbers in one’s head, a subject is highly likely to encounter the same information each time. We might call memory of this information “regeneration memory” to distinguish it from memory of the individual episodes. ... For the most part, regeneration memory provides a good source for determining how we did behave when the behavior concerns responses to invariant knowledge structures or stable aspects of the environment. (p. 164-165)

⁸ This phenomenon calls to mind Schmidt’s noticing hypothesis in second language acquisition studies, that all language learning is initially consciously noticed to some degree (Schmidt & Frota, 1986); the process only seems to be often unconscious because over time, the learner forgets the intermediate steps by which he or she learned.

Applied to the classroom context, it may be that a teacher cannot fully recall his thinking during a particular automatized behavior, but he may be able to fill in gaps by recalling previous memories of similar instances, in effect relying on memories of a routine (see 2.3.3).⁹

As to the issue of post hoc rationalizations, I think they do occur but it is extremely hard to discriminate between them and accurate recalls under non-experimental conditions (significantly, Nisbett & Wilson, 1977, drew heavily on experimental studies to reach their conclusions); given the aforementioned arguments in favor of the idea that many thought processes are conscious and may be recalled from memory, and given the general argument put forth by Ericsson & Simon—that whatever is noticed or heeded can be recalled—I think post hoc rationalizations may be the exception rather than the rule.

3.4.4 *Sampling and gaining access*

When I was deciding on my sample, theoretical and methodological considerations went hand in hand with practical considerations. I wanted EAP teachers because, as I discuss in chapters 1 and 2, most of the studies of teacher thinking and assessment have involved primary and secondary school teachers, and I wanted to look outside those contexts; an EAP context was a natural choice because it would be accessible to me (see next paragraph) and I had had several years' experience teaching EAP. I also wanted to involve more than one teacher teaching the same course; this was because I thought a multiple-case design would be much stronger than a single-case design. Yin (2003) explains that multiple-case designs allow a replication logic that single-case designs do not; the analysis can then be informed by similarities and differences across cases. Miles & Huberman (1994, p. 29), calling this “comparable case sampling,” includes this in their list of strategies that can bolster confidence in the analysis on the grounds of representativeness. In this study, I wanted to compare thinking of teachers who were given nearly the same course conditions (e.g., time.

⁹ Ericsson & Simon propose research designs that could differentiate whether participants were accessing episodic and regeneration memory, but they require experimental conditions; besides, it seems unnecessary to differentiate them if it is overall patterns that are being sought, as is the case in my study.

syllabus content, number of students, summative assessments), the primary difference being the teachers themselves.

Given these criteria, I approached a teacher at a university language center that was geographically conveniently accessible to me. I had met this teacher shortly after I had arrived in England for my doctoral studies, and he had expressed willingness to participate in any research I wanted to do. For one of my doctoral courses, I needed to conduct a classroom observation, so I contacted this teacher. He let me observe one of his classes and spared some time for an interview afterwards; I found him to be very helpful and quite articulate about his work, and we seemed to have established a rapport. Thus, I approached him again to ask if he could participate in my dissertation study; he was willing to do this. I also asked him if he knew of any other teachers who would be willing to participate, and he connected me with the second case study teacher.

The first time I met with Teacher 2 was briefly before the first class of his that I observed; I then talked with him at length afterwards. He too was very articulate and eager to discuss his work. I think his attitude and behavior—he was quite willing to spend time after his class for the recall sessions—would dispel any concerns that he was participating only out of a sense of duty to his manager Teacher 1, who was not only a colleague but also his manager (In fact, he was so available and willing to talk that I needed to make sure in my analysis that I did not give undue weight to his data!). He did make the request that I would let him copy the videotapes I made of his class sessions, because he worked a lot with multimedia and wanted them as a resource in case, for example, he wanted to provide teacher training (This did involve an ethical issue, as I discuss below in 3.6.). Considering what Teacher 2 was making available in terms of his time and energy, I thought this was a fair request and thus consented.

The methodological literature points out that within-case or internal sampling decisions about which activities, processes, times, etc. also need to be made. I discuss a few key internal sampling decisions below:

- *Course.* While Teacher 2 taught both an insessional EAP and a Pathway course,¹⁰ Teacher 1 taught only insessional EAP courses, so the insessional course was chosen to allow comparison between teachers.
- *Time.* This followed from the course decision. I observed Teacher 2's classes on Monday and Wednesday evenings, and Teacher 1's classes on Wednesday and Friday afternoons. Teacher 1 had other sections of the insessional, but the one I chose to observe was the most convenient in terms of time.
- *Location.* This followed from the previous two considerations. Both teachers' classes were held in the center's classrooms. Teacher 2 also regularly made use of the center's language laboratory classroom.
- *Class sessions.* I wanted to get a fairly complete picture of the teachers' assessment practices, so I attended nearly every class session. Actually, those assessment practices became fairly clear within the first few weeks (6-8 class sessions), but the extended time of observation and recalls allowed me to explore the cognitions and their sources more closely and to strengthen the reliability of my findings (through repetition).

An important internal sampling decision was which class segments to replay in the stimulated recalls. Based on my review of the literature (see 2.2.2 about assessment practices), I had some basic categories in mind: spontaneous assessment of student performance during a pedagogic activity, planned assessment of student performance, and tests. Thus, during each class observation, I noted periods of classroom action in which these general kinds of assessments occurred. In these two cases, the most frequent kinds were spontaneous and planned assessments; no tests were given by either teacher during the course. With spontaneous assessments, I included episodes when the teacher observed students doing an activity and episodes when the teacher questioned students. Planned assessments in these cases comprised mostly of arranged in-class student presentations, after which the teachers gave formative feedback. There were also occasional group discussions—this reflected one of the course's end-of-term summative assessments—for which the teachers also provided formative feedback. For the most part, when these segments of videotape were shown to the teachers for recall, the teachers did express that they were making

¹⁰ The intensive year-long Pathway course is for overseas students who wish to apply for a British university but want to improve their English before doing so.

judgments and gaining information about student language ability during those episodes, including moments of teacher questioning.

The preceding discussion has focused on describing and justifying the design and data collection decisions of Stage 1. The next section explains how analysis of the data proceeded.

3.5 Conduct of Stage 1 analysis

3.5.1 Summary of Stage 1 analysis

Table 3.2 summarizes the data sources and the tactics and practices used to analyze them.

Table 3.2 Overview of Stage 1 analysis

Data sources for analysis	
Teacher 1: field notes of 16 class observations; field notes & transcripts of 7 interview/recalls Teacher 2: field notes of 15 class observations; field notes & transcripts of 11 interview/recalls	
Generative analysis tactics	Confirmatory analysis tactics
<ul style="list-style-type: none">noting patterns and themesseeing plausibilitymaking contrasts and comparisonsclustering data & subsuming particulars into generalmaking metaphorsusing “hidden case”	<ul style="list-style-type: none">checking for representativenessweighting the evidencelooking for negative evidencetriangulatinggetting feedback from informants
Main analysis practices	
<ul style="list-style-type: none">using contact summary formskeeping a research diary on my computerusing visual devicescoding of transcript data using MAXqdadiscussing the ongoing research with colleagueswriting for external audiences	

As noted in Table 3.2, my data collection resulted in: for Teacher 1, field notes of 16 class observations plus field notes and transcripts of seven recall sessions; and for Teacher 2, field notes of 15 class observations plus field notes and transcripts of 11 recall sessions (I have included the interview data with the recalls because they occurred together.). To check the quality of the transcripts, I gave each informant

copies of their recall and interview transcripts and asked them to correct or clarify anything that was incorrect or liable to be misunderstood. This body of data was analyzed through a variety of tactics—some that generated findings and others that tested or confirmed those findings—that often occurred concretely within particular analysis practices. I now discuss how I conducted the analysis and issues arising therein.

3.5.2 *Generating findings*

In analyzing my data, I used several general tactics (the term used by Miles & Huberman, 1994) during and after data collection. Drawing from Miles & Huberman's list in chapter 10, the primary ones that I used to generate or produce initial provisional findings were, with a brief explanation, the following:

1) *noting patterns and themes*. In the process of looking over data, patterns or themes that draw together bits of data often emerge (Miles & Huberman, 1994; Punch, 1998). An important caveat to keep in mind with this tactic is that the mind sees patterns very quickly, so they need to be held somewhat loosely, with some skepticism, and be empirically checked.

2) *seeing plausibility*. Sometimes, an analyst makes a conclusion because it “makes good sense” or just seems to “fit” (Miles & Huberman). Like patterns and themes, plausible conclusions need to be supported or confirmed by other tactics.

3) *making contrasts and comparisons*. Having two cases in which the classroom aspects were mostly similar allowed me to examine what was similar and different about the two teachers' assessment practices and thinking from the outset.

4) *clustering data and subsuming particulars into the general*. To illustrate this tactic, here is an example from the Stage 1 data. When talking about their thoughts when assessing students, the teachers made statements like “it's important to give some positive feedback first, then the negative, and then end with a positive” or “when assessing students doing pedagogical activities, it's important to distribute attention fairly between students.” I grouped these together as “assessment principles”

(clustering), and since each teacher had a list of these principles that they referred to repeatedly, I subsumed these principles under the term “credo of assessment”—a personal set of principles that were kept in mind when assessing students.

5) *making metaphors*. Miles & Huberman (p. 250) write that this involves “comparing two things via their similarities and ignoring their differences. As Morgan (1980) notes, calling a boxer ‘a tiger in the ring’ evokes fierceness, grace, and power—and ignores striped fur, fangs, and four-leggedness. Metaphors are thus a ‘partial abstraction.’” My use of the term “projection” is an illustration of this. The teachers sometimes said that when assessing a student’s speech or writing, they would think about how the performance would be judged by, for example, native-speaker classmates during a class discussion, or by a native-speaker reader. I thought this resembled the psychological idea of projecting—or imagining or visualizing—oneself in a situation; hence my use of the term.

One tactic, or perhaps more accurately resource, that deserves special mention is using my own professional experience as a “hidden case”; as I analyzed the data, I often reflected on and compared it with my own teaching and assessing experiences. While I recognize the potential biases involved, I think in any kind of social research situation, there will be a balance that needs to be struck of being familiar enough with the subject being studied to have a working understanding of the issues involved and to relate to the participants, while at the same time being distant enough to maintain a kind of objectivity and to be open to new ideas and approaches to the subject. In any case, being aware that my own experiences could skew my analysis forced me to pay attention to check my findings (see 3.5.3 on confirmatory tactics).

I conceptualize the aforementioned tactics as thought processes that were ongoing throughout the duration of the study, but that often occurred especially in the context of specific practices, suggested from the qualitative data analysis literature, that I employed. Those practices included:

1) *using contact summary forms*. After each class observation and recall session, I typed out a one-page sheet that included a brief descriptive summary, impressions,

and issues to reflect on. This was then stapled like a cover sheet with the corresponding field notes and handouts. This was done to make referencing and access to the notes easier; it also helped me remember ideas and issues. An example of one such form I used is given in Appendix 4.

2) *keeping a research diary on my computer*. During the data collection period, I periodically recorded memos to myself, expressing issues that were being raised in the data collection process. The diary also provided a space to record, speculate and play with ideas; through this, I began to articulate analytic concepts and relationships. Finally, it provided an outlet for personal concerns and feelings. The excerpt below illustrates these:

5 Feb 04

¶ 1 The data collection went smoothly yesterday except that I had to run around to find VHS-C tapes! But [a colleague's] words still ring in my ears regarding "failing to prepare is preparing for failure." Actually, I think it's taken me the couple weeks to get sorted as to what I need to do, etc., and to take the time to type up schedules, etc. I guess once you get into a habit then it becomes like a routine, so it's not so overwhelming.

¶ 2 One thing I'm thinking about is describing these cases as examples where I can draw some kind of connection between macrolevel and microlevel assessment issues, between the social and political factors influencing teachers' assessment planning and context, and the online cognitive processes involved in assessment.

¶ 3 As I'm transcribing [CTB's] 1.28 tape, I have a thought: one way "the context" impacts assessment is on the time allowed for assessment. For example, with larger class size, there's less time for a detailed picture and only sketches can be drawn. So of course we know class size affects teaching, but here is evidence of a means: class size has a direct impact on the amount of time teachers have to listen to and to observe and to interact with (and thus formatively assess) students. ...

¶ 4 I'm taking a long time with transcription; am I doing something wrong? It took me about three hours just to transcribe about 30 minutes of tape!

One can see in this excerpt the genesis of two ideas that were developed and eventually made their way into the Stage 1 findings. One, in paragraph 2, was that wider social or political influences could impact teachers' cognitions and practices of assessment in the classroom (see 4.8.2), while the other was that classroom parameters such as class size influenced teacher thinking in assessment (see II under 4.5.1).

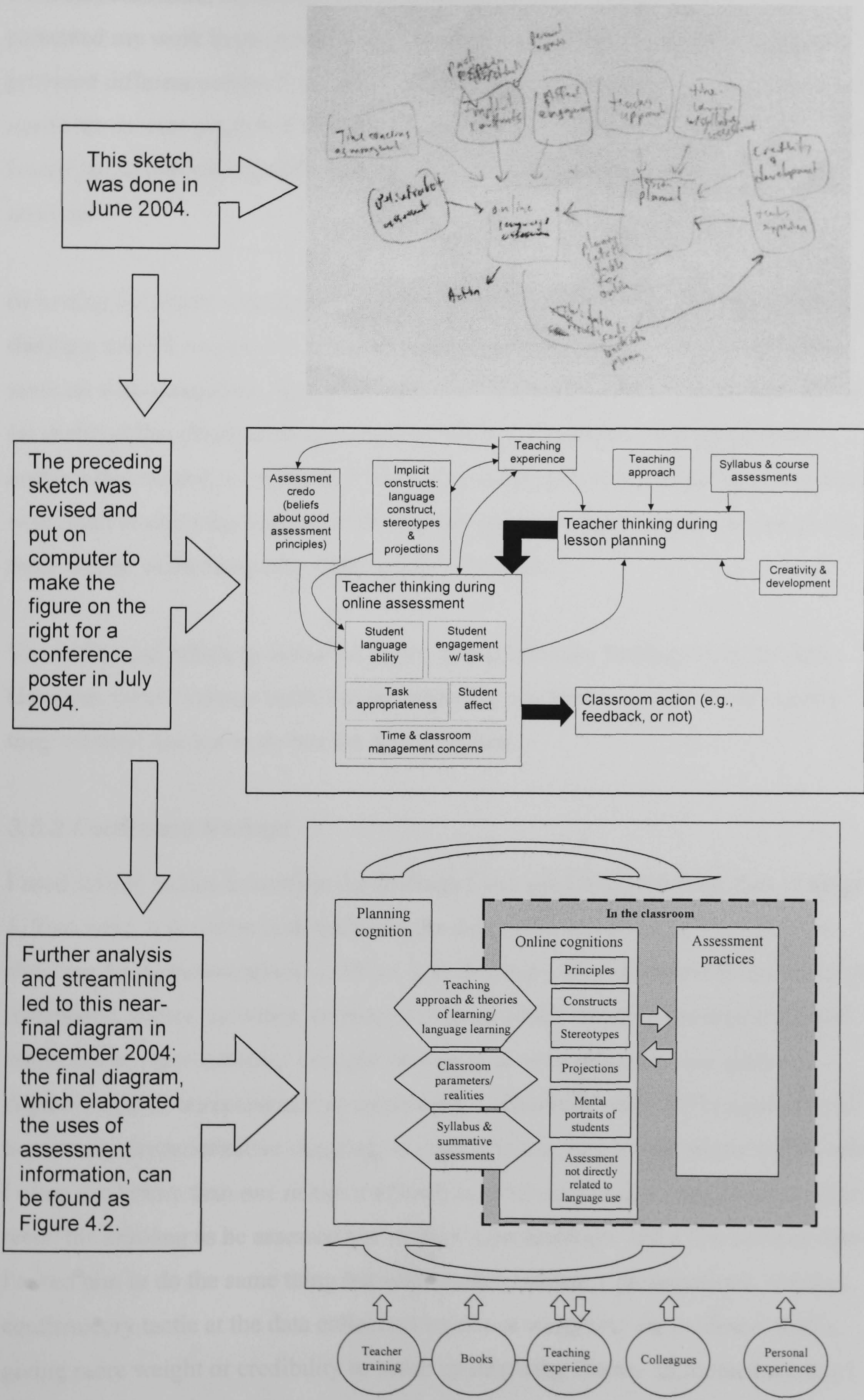
3) *using visual devices*. At a few points in the data collection and afterwards, I drew diagrams trying to describe relationships between the emerging concepts. This

helped me get a visual “map” of my analysis and allowed me to see gaps or data-thin areas that needed to be bolstered with further analysis and evidence. Figure 3.2 (see next page) shows the development of what eventually became the model of the assessment cognition network detailed in chapter 4. Interestingly, I believe the desire for an increasingly more “aesthetic” diagram also led me to check relationships and categories for their connections and relevance.

4) *using the qualitative data analysis software program MAXqda* (Kuckartz, 2001).

This program aided analysis in several significant ways (see Appendix 5 for an annotated screenshot). First, I used it to code the transcript data. The general process of qualitative coding has been covered in detail in the methodology literature, so I will only describe it in general here. The process essentially involves examining the data for topics, patterns, issues, etc.—using tactics like the ones described above and/or drawing ideas from the literature—to generate representative words or phrases (“codes”) to which relevant excerpts of data (and sub-codes) are assigned. While this is the basic process, I found that there was much going back and forth between codes and data, with many codes being generated, revised, or removed to match the data more closely; this was particularly true when I applied the set of codes generated from the first teacher’s data (which I had done first) to the second teacher’s. The overall goal of the coding process was to ground analytic concepts in the data, providing a structured and organized basis from which to draw conclusions. Second, MAXqda enabled me to insert memos alongside the codes. These memos were places where I could record ideas and thoughts that arose in the process of coding or that I had gleaned over the course of other analysis practices (like from my research diary) and wished to connect to concrete data. Finally, the program facilitated tactics for confirming or testing my findings (see 3.5.3).

Figure 3.2 Iterations of the assessment cognition network



5) *discussing the ongoing research with colleagues*. This includes talking with my doctoral classmates, my supervisor, and participants at conferences in which I presented my work in progress. These conversations were invaluable because they provided different perspectives and feedback; I know at least two instances in which useful terms were generated from these interactions. Talking about my research also forced me to sharpen and articulate more clearly what I was finding through my analysis.

6) *writing for external audiences*. As has been recognized (Wolcott, 1990), writing is thinking, and I found myself doing a great deal of analysis as I wrote for feedback sessions with classmates, for a presentation or a poster I gave at two conferences, or for drafts of this dissertation. As I wrote, I sensed where there were gaps in my analysis that needed to be filled, or possible validity concerns that needed to be dealt with, such as checking findings with multiple instances from both cases, and giving more weight to higher quality data (see 3.5.3 below).

The tactics and practices described above generated many findings from the data. However, these findings needed to be verified or confirmed in order to strengthen their validity. Tactics to do this are described next.

3.5.3 Confirming findings

I used several tactics to confirm the findings I was generating from the data in Stage 1. Two were, it could be said, built into the data collection process. The first was *checking for representativeness*. Miles & Huberman (1994) point out that samples of informants, events, activities, or processes upon which tentative findings are based might not be representative. I sought two cases so as to safeguard—at least to a degree—against unrepresentative informants, I observed nearly all the classes so as to prevent unrepresentative sampling of class sessions, and for the stimulated recalls I often used more than one instance of each assessment practice (e.g., I had a teacher recall his thinking as he assessed one student's presentation, and a few sessions later I asked him to do the same thing but with another student's presentation). Another confirmatory tactic at the data collection level was *weighting the evidence*: that is, giving more weight or credibility to better quality data. For the stimulated recalls, I

knew that the longer the time between the original event and the recall, the more likely the effect of memory decay (see III.2b under 3.4.3). Therefore, I tended to ask informants to recall more recent assessment episodes rather than earlier ones; in analysis, I then tried to give more weight to those later episodes.

I also verified findings by *looking for negative evidence*. Once I had a plausible finding, I tried to revisit the data for counterevidence. For example, a couple of CTA's early comments made me think that he held a kind of "technicist" ideology in highly valuing quantification and efficiency in assessment [e.g., A1:41-42]. However, as I looked at later transcripts and my observation data, it was clear he did not use quantification in his classroom assessment, and I realized upon re-examining those earlier comments that CTA had made them in the context of reporting summative assessments.

Triangulating, in which findings are supported by independent means, was also another confirmatory tactic. Denzin in Miles & Huberman (1994) distinguishes several forms of triangulation, including by different data sources, methods, and researchers. In terms of data sources, I checked findings across the two cases and across recall sessions within each case; in terms of methods, my use of observation and recalls provided some degree of triangulation, although strictly speaking the recalls were not independent of the observations. These ways of triangulation were also bolstered by the data sources and methods of Stage 2 (see 5.2 and 5.3.1).

I did not triangulate by having another researcher look at my Stage 1 findings; instead, I used another tactic, *getting feedback from informants*. After I had written up a draft of the Stage 1 findings, I sent slightly shortened versions to the two teachers for comment. Both teachers gave positive responses; one replied, "I've just finished reading your draft and everything seems fine to me—an accurate reporting of the data and a rational, scholarly interpretation of what you've researched." and the other wrote, "I found it entirely well written and argued and a very stimulating read." [e-mails from 14-15 June 2005¹¹]

¹¹ The late date of these e-mails reflects that 1) I did not write up the Stage 1 findings until early 2005, and 2) I recognized as I was writing my dissertation that I needed to bolster the validity of my findings (see 3.6).

3.5.4 Concluding comments about Stage 1 analysis

I have devoted some space to describing the conduct of my analysis primarily because while analytic techniques for quantitative data are widely agreed on and formalized, analysis with qualitative research has often been seen as the “black box” in which data goes in and are magically transformed into findings (Miles & Huberman, 1994); in framing the analytic process in terms of cognitions and practices,¹² I wish to highlight that, on one hand, qualitative analysis involves a complex set of processes, and yet, on the other hand, those processes can be made relatively transparent to research audiences.

I saw the findings of my analysis of the Stage 1 data, presented in Chapter 4, as generally exploratory; as I showed in Chapter 2, there has been relatively little work done in the area of teacher thinking in relation to classroom assessment, with even fewer studies of such thinking in naturalistic contexts. However, as it turned out, there was a great deal of similarity between my Stage 1 findings and the literature, particularly in regards to the considerations—such as beliefs about teaching and learning, and classroom parameters—that impinge upon teacher thinking in assessment. In a sense, then, my Stage 1 findings were confirmatory of findings from other researchers’ initial forays into this topic (see 7.4). Meanwhile, some of my findings seemed to be unique and not mentioned in the literature, such as the impact of organizational policies and the use of stereotypes based on nationality when assessing students; they needed confirming. The analysis also raised further more specific issues beyond my original set of research questions that I thought demanded exploration, like the question of how to improve teachers’ impressionistic knowledge of students. It was these findings and issues that were explored in Stage 2, discussed in chapter 5.

3.6 Ethical issues arising in Stage 1

As Kimmel (1988) points out, an ethical dimension underlies any research endeavor. even from the start, where the researcher’s values play a role in the way a problem is framed. Generally speaking, I attempted to maintain an “ethical mindfulness” (Bond, 2000)—a sensitivity to possible issues—throughout my conduct of the study and to

¹² The resemblance to the way I have framed teacher thinking and assessment practices is obvious and probably had much influence in my thinking while writing this section.

allay potential ethical problems; so, for example, besides the usual step of asking for their consent to participate (see Appendix 6 for all consent forms used in this study). I tried to make as clear as possible to potential participants what I was planning to do and what kind of commitment they would need to make. However, as it turned out, each stage presented ethical issues that I had not expected and that were particularly problematic. Most of these issues arose in Stage 1 and are discussed below: a dilemma arising in Stage 2 is discussed in 5.4.

The first issue involved videotaping and informed consent. In my consent form to students in one of the teachers' classes, I stated that I would be videotaping the classroom; I also mentioned that the data could be used for research and training purposes, but I would observe standard anonymity practices like changing names. Besides the fact that I had not realized that using videotape for training purposes would violate anonymity, another dilemma arose: on one hand, the teacher had given his consent to be videotaped; on the other hand, a few of the students said they did not consent to being videotaped. In the end, I said I would try my best to keep students who had expressed non-consent out of the picture frame, although I could not guarantee this. Most were satisfied, although one—a student from Afghanistan—was noticeably uncomfortable when the camera was turned in the student's direction, sometimes putting a hand up to block out the line of sight or leaning far back in the seat. As mentioned earlier, a couple of students also regularly sat with their backs to the camera. If I were to do it again, I would likely state that students not wanting to be videotaped would be blurred or pixellated if the videotape was used for training or presentation.

A related issue involved making something of a deal with Teacher 2 as part of gaining access. As I talked with him early in the term, he requested that I lend him the videotapes so that he could copy them for his own use. As I mentioned earlier (see 3.4.4), I thought this was a reasonable request, especially considering that he was eagerly giving of his time.¹³ I later realized that this was not part of the consent form I had given to students. I then told the teacher that I was quite willing to share

¹³ In addition, he lent me a VHS-C to VHS converter, making the stimulated recall sessions much more convenient and timely, as I only had a VHS-C camera and the videoplayers in the language center were all VHS.

the videotapes, but added that I hoped he would get the consent of students if he did use them. However, this was somewhat a token statement, as I did not demand it of him. Technically speaking, the consent forms did say the data could be used for training purposes, which was what the teacher planned to use the tapes for, but I think the consent form should have anticipated this possibility; at the least, I should have asked the teacher to gain the students' consent again with his own form.

The third issue related to taping of summative assessments. I had originally planned to include stimulated recall data from the course's summative assessments, particularly the group discussion task. On the day of the assessment, I taped the first group discussion, which involved five students. The problem was that one of the students was not in my case teacher's class (the groups consisted of students from different teachers' classes); he had not consented to my taping, and I was concerned that my taping had had a chilling effect, as he talked least among the group participants. I did not tape any others, partly because I had second thoughts about doing so, and while I did conduct a recall session with the case study teacher, I decided not to include it in my analysis, and I did not tape any more group assessments. I think this was an ethical misstep that could have been prevented by being careful to tape only a group whose members had given me consent, or by deciding at an earlier stage not to include summative assessment data because of the potential negative impact on students.

In this section, I have tried to be candid about the ethical difficulties I faced while conducting Stage 1 of this investigation. I have done so in the interest of research validity, which can be seen as an ongoing ethical concern in itself (Miles & Huberman, 1994). I have not explicitly discussed validity in this chapter, but I hope it is clear that in fact a concern for validity—accuracy and trustworthiness of the data collection and analysis—underpins both of my methodology chapters (chapters 3 and 5). Besides describing the data collection and analysis processes I undertook, I am also arguing for the validity of each level—method, strategy, and stage—of my study: in effect, I am arguing for the overall validity of my methodology by arguing for the validity of the parts that comprise it. In fact, my purpose in describing my data collection and analysis is to make the process transparent and “auditable,” thus strengthening reliability in the qualitative research sense (Miles & Huberman, 1994)

and therefore adding another argument for the validity of the methodology, and by extension the validity of my findings.

3.7 Summary

In this chapter, I have set forth the paradigm from which I am working, provided an overview of this study, and outlined the principles underlying its overall design. In addition, I have explained the decisions in my data collection design and execution, articulated my analysis procedures, and disclosed ethical dilemmas arising in Stage 1. I follow a similar pattern in chapter 5 when discussing my Stage 2 methodology.

The findings from Stage 1 and Stage 2 are given in the next chapter and chapter 6, respectively.

Chapter 4 Stage 1 Findings

4.1 Introduction

In the previous chapter, I gave an overview of the study's research design and outlined the methodology of Stage 1. In this chapter, I present the findings from that first stage of the investigation.

Below, after initial background information on the teachers and the course they taught, the cases are presented together according to the research questions listed in 3.4.1 (rather than one complete case after the other). I have chosen to do so partly for efficiency of presentation, but mainly because the cross-case comparative analysis (Miles & Huberman, 1994; Yin, 2003) was much richer than the within-case analysis alone. In addition, this reflects the data collection; I conducted the two case studies concurrently, and so comparison was an analytic technique from the beginning. In answer to each research question, I explain how the data was elicited and analyzed, then describe the findings; I also discuss several of the findings in relation to other findings from this study and in relation to the research literature. Data are referenced by [Teacher TranscriptNumber: LineNumber] (e.g., [A1:1]); where there is no accompanying data excerpt, such references are provided to show where data evidence is located, for purposes such as data audit and researcher reference. In data excerpts, *** stands for unclear speech that could not be transcribed, while S and another letter (e.g., SA) stands for a student and M stands for the researcher.

4.2 Background to the cases

The participants in Stage 1 were two teachers at a university language center in the UK. CTA,¹ a male native speaker of English, had taught English (and sometimes French) for over 20 years in primary, secondary, tertiary, and private language schools both domestic and abroad. He had been teaching at this center full-time for six years, and had added the managerial role of EAP coordinator a year before I observed him. He had not gone through extensive initial teacher training, but later in his career did complete an RSA Diploma in TEFLA,² which he felt was “one of the

¹ As mentioned in chapter 3 note 4, for the sake of anonymity, I use the abbreviations CTA and CTB to refer to Case Teacher A and B. In transcripts, I use A and B, respectively.

² Royal Society of the Arts Diploma in Teaching English as a Foreign Language to Adults, a teaching qualification replaced by the Diploma in English Language Teaching to Adults (DELTA).

best things that happened to me; I really felt I was a professional for the first time in my life.” [A6:68]

CTB, also a male native speaker of English, had been teaching English for over seven years, mainly at this language center. Having worked in a variety of fields before teaching, CTB completed a CELTA³ course in 1996, and began work at the center that same year; he had continued at that position up to the time of the study. During that period, he also completed the DELTA, a more advanced qualification after the CELTA, in 2000.

In terms of their organizational relationship, CTB and a few other teachers reported to CTA, who in turn had the English programs manager EPM over him. Based on observation and their comments, CTA and CTB had an amiable working relationship; decisions about syllabus and assessments for EAP courses were CTA’s responsibility (subject to EPM’s approval), but he took the views of CTB and other teachers into consideration, too. More collegially, they also shared classroom activity ideas and resources.

Both of these teachers were teaching the same course, an insessional EAP course. The teachers often compared this course with the center’s preessional course—the insessional drew much of its content and the structure of its summative assessments from the preessional—so both are described below:

- The center ran intensive preessional EAP classes—that is, classes before the beginning of the academic year—in the summer, mainly for overseas students who needed to fulfill a language requirement before being fully accepted for their undergraduate or postgraduate program. These lasted from five to 12 weeks, with students attending 15-20 hours of class a week.
- Insessional classes, on the other hand, were run during almost every academic term, for a more diverse group of students. They included undergraduates on the SOCRATES/ERASMUS program,⁴ undergraduates on a special university-sponsored program for overseas students, and postgraduate overseas students.

³ Certificate in English Language Teaching to Adults, an initial teaching qualification given by Cambridge ESOL.

⁴ SOCRATES ERASMUS is an European Union student exchange program in which students from different countries can study for part of their undergraduate degree in another country.

This was complicated by the fact that some students had to take the insessional course as a credit-bearing requirement while others audited it. The students were also diverse in incoming language ability, because students were placed into particular classes based not upon placement testing but upon when students had time in their class schedules. Students attended two 90-minute sessions a week for the 10 weeks of a term.

The content of the insessional course focused on English use in academic contexts (see Appendix 7 for the syllabus and scheme of work) and could be seen as a reduced version of the preessional course. As Appendix 7 shows (and classroom observations confirmed), students generally spent one session a week on academic reading and writing and the other session on listening and speaking.

The assessments that were observed—and thus the data covered below—happened within the context of pedagogy; the information gained from such assessments were used for classroom purposes, although as I stated in 2.2.2 my definition of assessment also allows for the collection and interpretation of information without an explicit use. There were only two summative assessments for this course, and these occurred a few weeks after the last class (because of the between-term holiday): 1) students had to read three texts and then write an argumentative essay drawing upon the texts as sources; 2) students had to watch a video segment on a topic and then participate in a group discussion with 3-4 other students.

These two summative assessments had been put in place relatively recently on the preessional course, and were being applied to the insessional course. This was part of a wider “formalization” process being implemented by the management (see general remarks about formalization and managerialism in 1.2). A few years earlier, there had been no unified assessments; teachers taught what they wanted and made their own exams [A3:106]. However, changes in the management structure led to the hiring of EPM, who had previously been at a language program of another university, to coordinate the center’s English programs, which included business and Pathway English in addition to EAP. EPM then gave CTA, who had been one of the teachers, the additional role of coordinating the EAP courses. Under EPM’s direction, both the syllabus and the summative assessments for English programs—particularly the

preessionals—had become more explicitly specified and made more uniform across teachers. One reason for this with the preessional, according to CTA, was that EPM had taken on the task of convincing other departments in the university that the language center's marks for students on preessionals were as or more valid than Cambridge IELTS scores [A3:98].⁵ This change for the preessional in turn influenced the insessional class, since the content was largely based on the preessional. Another reason was that EPM wanted the course assessments to match what the departments required of students, like giving presentations, participating in class discussions, and writing essays [A7:82]. Finally, the insessional was a credit-bearing course, which meant such explication and standardization was imperative because it would be externally examined [A3:100-106].⁶ For the insessional course involved with the case studies, CTA had been responsible for specifying the syllabus and the assessments, with input both from EPM and teachers of the course, including CTB.

In terms of class composition during the term observed, there were consistently at least 12-14 students in CTA's class and 6-8 students in CTB's class; most of these students enrolled at the start of the term, although a few enrolled later. The students in both classes were either ERASMUS/SOCRATES or postgraduate students, with an occasional undergraduate student attending CTA's class. Also, CTA's class consisted mostly of women, while CTB's had slightly more men. Importantly, the students' English language levels varied, because students enrolled on inessionals based on how classes fit into their timetable, not based on any kind of language placement scheme.

Given this background, answers to the research questions posed in chapter 3 are presented below.

⁵ Arguments included the fact that students on preessionals did work similar to required work in their future departments, and that the marks were based on assessments over time rather than a one-off mark like the IELTS.

⁶ In the UK, programs offered by university departments must undergo regular external examination, in which examiners—usually academics from similar programs at other universities—evaluate various aspects of the program, such as course content and assessment arrangements.

4.3 RQ1: What are teachers' definitions of assessment?

I was initially very concerned about the possible gap that may have existed between my definition of assessment (see 2.2.2) and the teachers' definitions. As Kennedy (1997) notes, one significant barrier to communication between researchers and practitioners is different conceptualizations of problems and concerns. Kramsch (1995) has also pointed out that researchers and practitioners form two communities that have different ways of discourse. However, in the event, my and the teachers' views did overlap and we were able to communicate clearly.

CTA saw assessment as helping students, with feedback forming an important aspect.

I'm giving them feedback on their learning, so that they know whether they're learning or not. Well, learning language is a skill, so I'm telling them how they are advancing in their skills, their language skills, doing assessment, whether it's feedback on a presentation, or a ... progress test, or final assessment. [A1:13]

CTA also saw assessment occurring both formally and informally:

Informal is as looking over their shoulder and saying "yes that's good," or "no you need to do this" or "your pronunciation wasn't quite right there, it should be this." ... So it could be error correction, that's a form of assessment, really. Or not, that isn't in itself, but ... a comment: "that's fine," or "that's not quite right," that's an assessment, isn't it, that's an evaluation, that counts as assessment, that's the most informal. Formal are progress tests and the final assessment and also external examinations, an IELTS exam for example. [A1:15]

When first asked to define assessment, CTB responded with a metaphor of assessment as, ultimately, helping students move along the journey towards their destination, e.g., entrance to postgraduate study:

So what is assessment very much depends on what direction ... the students are moving in, in order to work out what stages or what posts they will have [to get] past on that journey to be in a position to take that next step... . [B1:5]

Hence, like CTA, CTB saw assessment as a key to student development. Also like his colleague, CTB differentiated between formal and informal assessment. Formal assessment was any procedure in which the teacher's assessment of a student's work had an impact on whether the student succeeded or failed on the course. Informal assessment, on the other hand, "doesn't carry the threatening weight of deciding

whether they succeed or fail on the course or not.” He added, “And informal assessment would be just the way I’m observing and collecting evidence about each student or forming my general opinion about their skills and their abilities.” [B10:40]

As the above data shows, where the teachers’ definitions significantly differed from the one I give in 2.2.2 was in that they saw all assessment as ultimately aimed at developing students, rather than as aimed at teacher decision-making (although one could argue that decision-making is itself ultimately aimed at developing students). On the other hand, the teacher’s definitions shared some important aspects with mine. First, while they did not emphasize cognition as my definition does, CTA’s comment about “looking over their shoulder” and CTB’s comment about observing and collecting evidence implies a cognitive dimension to their definition. Second, they focused on student ability, which aligns with language use in my definition—although CTB does not explicitly state linguistic ability, and in fact looked at critical thinking, too (see II under 4.5.2). Incidentally, these responses differ from those found among primary and secondary teachers, who emphasized not only the assessment of ability but also effort (Mavrommatis, 1997; McMillan, 2003); the case teachers rarely mentioned student effort. Presumably, this is because teachers become less concerned with student effort as student maturity and schooling socialization, and perhaps teacher recognition that older students have “more things going on in their lives”, increase.

A third shared aspect of the teachers’ and my definitions—arguably the most important in terms of methodology—was a view of assessment as more than just formal testing. Most notably, they included the day-to-day observation of students, which meant that my operationalization of “assessment practices” to include excerpts of classroom activity that seemed to involve observation of students also made sense to the teachers.

4.4 RQ2: What are the classroom assessment practices of EAP teachers?

Table 4.1 summarizes the two teachers’ classroom assessment practices in the insessional course. These lists are based on classroom observation field notes and

videotaped data (see I under 3.4.3 and Appendix 2), with a practice being operationalized as any pattern of action in which the teacher may gain information about student language use; the practices listed here can be seen as specific forms of the general practices described in 2.2.2.

Table 4.1 Classroom assessment practices of teachers of an insessional EAP course, listed in order of prominence/frequency

CTA	CTB
1) in front of the whole class, calling on students and asking (often closed) questions [11.2.04]* 2) observing group work on planned tasks [28.1.04] 3) observing individual speech performances [13.2.04] 4) observing structured group discussions [3.3.04] 5) observing group written texts [17.3.04] CTA also gave feedback on written homework (not included in this study).	1) observing group work on planned tasks [4.2.04] 2) in front of the whole class, asking fairly open questions for anyone in the class to answer [2.2.04] 3) observing individual speech performances [16.2.04] 4) observing structured group discussions [8.3.04] 5) observing group written texts [18.2.04] CTB also gave feedback on written homework (not included in this study).
* Example class dates during which the practices could be found are in brackets; often, a few practices were used in the same class period, but I have used different dates for each to show that these practices were used throughout the term.	

The table shows that while the two case teachers shared some assessment practices in common—observing individual students’ speeches, group discussions, and group written texts—there was a significant difference in their plenary time questioning and observation of group work. CTA spent a good deal of class time with the whole class’ attention, calling on individual students by name and asking them questions that often had right and wrong answers (i.e., closed questions). He also set activities for the students to work on, often in pairs or groups, and observed them, but he did not spend nearly as much time on this as CTB. CTB usually set up activities for the students to work on, also in pairs or groups, and then spent most of the class time going around to each group to observe, comment, and answer questions. He did have the whole class’ attention at several times during a session, but did not spend as much time in this situation as CTA. In addition, during those times, CTB usually asked open-ended questions that were not directed at a particular student but could be answered by anyone. To illustrate this difference in questioning, excerpts of classroom discourse from each teacher are provided in Appendix 8.

The differences described above in the two teachers’ “styles” of assessment practices can be mostly attributed to the teachers’ differing teaching approaches, as explained in I under 4.5.1.

In fact, based on the classroom observations, it seemed that assessment in CTA’s class was often more “public”—occurring in front of the whole class—than assessment in CTB’s class. The “public” vs. “private” types of assessments can be seen on a continuum, as in Figure 4.1:

Figure 4.1 A continuum of assessment “audiences”

public	semi-public	private
calling on a student during plenary discussion	observing a student working in a group of 4-5 students	observing a student working individually or in a dyad

In Figure 4.1, the assessment practices given are placed from left to right according to the size of the audience; thus, calling on a student during whole-class discussion is more public than observing and giving feedback to a student in front of several classmates, which in turn is more public than doing so to a student working alone or with a partner. This differentiation between public and private is significant for at least two reasons. First, in the data, it seemed CTA was sometimes concerned about not embarrassing a student when observing and then giving feedback (see I under 4.5.2 and Table 4.2); this was mentioned during recalls of public and semi-public [A4:18; A5:13] assessments. It may be possible that concerns about “face” and avoiding embarrassment for a student impinge more on teacher thinking as assessment becomes more public, thus perhaps affecting judgments or feedback given. Second, there may be a time control issue involved; when time needs to be managed closely by the teacher, this may lead to more public assessment as the teacher takes more control (see further discussion under 4.8.2).

Having described the teachers’ assessment practices, I now turn to the thought processes that seemed to accompany them.

⁷ I wish to thank my fellow doctoral student Wen-Ding Huang for these terms.

4.5 RQ3: What cognitions underlie these practices?

The following findings result from my analysis of the stimulated recall and interview data (see 3.5). As I mentioned in 1.2, I use the term “cognitions” to include all the thought processes that teachers draw upon during assessment. The ones found thus far are organized under two headings:

1) strategic⁸ cognitions: teaching approach and beliefs about language learning, classroom parameters, and course syllabus and summative assessments. These substantially influenced teacher thinking in relation to assessment, mostly during planning but also during class time.

2) interactive cognitions: deployed during interactive assessment, these were assessment principles, constructs applied interactively, stereotyping, projection, mental portraits of students, and assessment not directly related to language use. These were operative mainly as teachers assessed students during class time, although they were also used in planning.

I also discuss the uses to which teachers applied assessment information. In brief, while the information was used mostly for classroom purposes such as feedback and management, it was also found that the information could inform summative assessments of students.

These cognitions and uses are elaborated next.

⁸ I wish to thank one of my examiners for suggesting this term.

4.5.1 Strategic cognitions

I. Teaching approach and beliefs about language learning

As Table 4.1 shows, the teachers displayed individual profiles in terms of frequency and quality of assessment practices. This seemed to be based largely on their language teaching approach and beliefs about language learning.

When asked about his prominent practice of calling on students, CTA said he had done something like it in his primary and secondary teaching, but it really became intentional and then automatic when he went through his RSA diploma training. In that training, he was taught a presentation-practice-production methodology by a particular trainer; the trainer explicitly advised him to call on students by name, to elicit a response [A4:6-8,25-27]. Calling on students also accorded with two beliefs about language learning that CTA expressed. One was explicitly connected: “I think [calling on students is] trying to stretch students, and this idea of stretched output; you’re getting students talking when they don’t feel they can, because that’s when they’re learning.” [A4:14] The other was expressed in the context of CTA explaining why he gave feedback. “... From what I’ve read and think about it, a lot of language learning is receptive. They’re probably learning as much language by listening to me as they are by speaking in class.” [A6:31] While CTA did not connect this directly to calling on students, it seems that this belief about learning from listening accords with spending time in “public” assessments. CTA also believed that interaction was important for language learning (Ellis, 1994)—hence group work—although it was not so easy to assess students when the whole class was in the midst of interactions [A2:51].

When asked about his prominent practice of setting up group work and then observing them in that context, CTB ascribed it to

discovery-based communicative language teaching: put the students in a situation where they’re discovering something, see what they bring to you and respond to what they’re ***ing, as opposed to laying it out really clearly structured. ... you throw them into a situation with the language and give them a task to do, and then you respond to what they come up with. [B8:27,29]

CTB also emphasized the importance of having students exchange information through, for example, information gap activities [B2:4]. As with CTA, this reflected a belief in the importance of interaction in language learning.

II. Class parameters

The givens of a class—especially the number of students and the time and timing of class sessions—were significant in the case teachers’ assessment thinking. The following excerpts show how awareness of classroom parameters played a role when teachers assessed students in class, particularly in the amount of time spent assessing each student and the quality of the feedback given to students:

A: I have to tread very lightly with a class that I only meet three hours a week. If they were full-time intensive [i.e., a preessional], I would be looking much more closely at their work; I’d be scrutinizing them more. [In this class, I am] just trying to pick up on the main features. ... I can’t show them all their weaknesses, because I haven’t got the time to put everything right. I just try and get the worst ones, because otherwise they’d be disheartened if I gave a whole list of errors, every detailed error, and said, “OK, see you on next Wednesday...,” whereas if they’ve had them full-time, I’d say, “look, right, this next lesson we’ll go through the whole list.” I haven’t more time, so I just try and pick up on the main problems, or even more, just getting them to look at the problems themselves; go down to the self-access centre, use the grammar books. I’m also aware that many of them aren’t really up to the university’s minimum requirement, I would say, of 6.5 IELTS or the TOEFL equivalent of 600. They’re allowed to squeeze in, but they’re here now so I don’t want to embarrass them, or worry them unduly. [A7:7]

M: So do you think that—getting back to an issue we talked about last time—if this had been a larger class...

B: [Assessing would be] a lot harder.

M: Do you think you would have still tried to put something like [spending time with each group] into effect?

B: Yeah. I tend to, I have to move around and take much quicker snapshots, so I’m sitting there for quite a while between these two groups and I’m following quite a lot of their conversation. With an inessional group with 16 people, which happens quite often, ... there might be groups of two or three, and I’ll basically be spending 30 seconds with each group and just rotating around. And I’ll get very sketchy impressions. ... To be honest, if this group is going to stay this small, I’ll start to feel that I can develop quite a good awareness of the individual students, where they are and their needs, and to some extent tailor my responses to each of them to suggest directions they can move in and respond to their needs. If the group gets larger, I tend to take a step back and realize that I’m not realistically going to be able to do that. [B2:21; also B3:47]

It seems this awareness of classroom parameters can figure into both planning and interactive thinking. For example, when planning, a teacher would consider how to make time for all his students to do presentations (which would be assessed for feedback) during the term. During class, the teacher would adjust plans and make online decisions taking into account the actual number of students present.

III. Course syllabus and summative assessments

As mentioned in the background to the cases (see 4.2), there had been increased managerial action in recent years; most notably, 1) a new layer of management had been added to the organizational structure, and 2) the syllabus and summative assessment activities had become more specified and uniform. The second especially had a considerable impact.

First, the syllabus and summative assessments were both important in planning. Both CTA and CTB implied or stated that they considered syllabus content when planning activities for a class session [A7:44; B4:21]. In addition, the forms of the summative assessments (an essay and a group discussion) were used in class for formative purposes [3.3.2004 observation of CTA; 8.3.2004 observation of CTB] and for preparation [A7:18; B11:22].

Second, the syllabus and summative assessments played a few roles during interactive thinking. The syllabus content comprised part of what the teachers looked for in interactive assessment (see discussion of constructs in II under 4.5.2). Also, the specific nature of this syllabus and amount of coverage it demanded impinged upon interactive thinking indirectly in the form of time pressure—this was particularly a problem for CTB; for example, in one instance he had to decide whether to cut short a student presentation because of time concerns about getting through the syllabus content [B5:18]. The summative assessments also occasionally influenced thinking, at least in CTB's data. This happened through either 1) similar processes, like below:

Basically, if you take the descriptions of what I said I was doing with SN and what I said I was doing with SM and what I said I was doing with SC, I would make all the same observations; I'm thinking about what preessionals do when they give an assessed presentation, I'm thinking about what I would be noting down and comparing with my colleague in assessing that, so it's pronunciation features, features of the presentation,

such as the aids used and like I said the communicative demeanour towards the group, grammatical and structural features, what features of the performance actually interfere with the listener's understanding, and so if you heard the things that I wrote down for SY there, it's basically the same set of feedback areas that I gave to SN and SM. [B7:5]

or 2) through the application of summative “products” (here, bands) to interactive assessment:

When I'm listening to the others I'm confirming [previous judgments about their language use], and I was looking forward to the discussion assessment at the end and monitoring the students in that situation—how well they did, how well have they organized their thoughts—and my basically feeling was again SJ was probably... he seemed to have quite a bit of difficulty forming an argument, sometimes it's hard to judge, but it may be the question that he had chosen. But even so, it's not such a worry as to imagine that he's going to drop below the class; he may be in the lowest band or he may be in the second band up, so there's nobody that I'm really worried about in the spoken aspect. [B3:53]

As mentioned at the beginning of this section, the above issues were labeled strategic cognitions because they had a wide and substantial impact on teacher thinking in relation to assessment, influencing cognition in both lesson planning and class time. It seems the way they did so was as follows. First, when the teachers planned their classes, they seemed to have in mind an array of elements that they worked with: beliefs about language teaching and learning (and accompanying methods), knowledge of classroom parameters, and syllabus and assessment requirements; they also had in mind some previous knowledge of the students from assessments in previous class sessions [A1:62]. Planning then “set the table” for subsequent assessment; that is, the lesson plan set the situational parameters (e.g., whole class vs. small groups, closed vs. open tasks, task content, etc.) within which interactive assessment would occur.⁹ Thus, one could say that these strategic cognitions affected interactive assessment mainly indirectly, although as the previous data shows, it could be direct, too.

It is specifically to other, more direct influences on interactive assessment that I now turn.

⁹ A similar idea, that certain “micro” decisions are nested in earlier “macro” decisions, is expressed by Woods (1996).

4.5.2 Interactive cognitions

The following cognitions were mentioned predominantly in relation to interactive assessment; that is, when the teacher was making immediate judgments about students during class time in the context of observing students working individually or in groups, interacting with them in class discussion, or observing students give a performance like a speech or a group discussion. It should be noted, though, that the cognitions below were not used “purely” in interactive assessment; they could also influence planning. For example, CTA had a stereotype that German students taking the insessional class had relatively strong language abilities and thus planned to pay somewhat more attention to weaker, non-German students (at least for one particular class session) [A5:28-39].

I. Assessment principles

As I examined the interview and stimulated recall data, it became apparent that the teachers referred to certain assessment principles or maxims; such principles were referred to both as they talked about assessment generally and about interactive assessment particularly, as the excerpts below illustrate:

Give positive, negative, then positive feedback

[CTA is describing his thinking as he observed and gave feedback to two discussion groups.] Trying to sandwich the feedback. Something positive, something which needs to be worked on, followed by something positive. Always works pretty well. [A3:52; see also A1:31]

Do not demoralize students with too much negative feedback at once

[CTB is describing his actions when marking written work.] B: And then also when you're marking somebody's piece of work, you can be a bit selective, so if a student's work is really strange English because the L1 problems are so strong, what you'll do is you will just pick the most significant errors and you let a lot of other things pass.

M: To prioritize.

B: Yes, because you don't want to demoralize them too much, so you want to draw attention to what needs patching up first, which means that usually my strategy is to go through and put the question marks first, because that's the place where the student isn't communicating, and if the student isn't communicating, that's the main problem. [B10:29-31]

Each teacher had a personal set (or “credo”¹⁰) of principles related to assessment. These are listed in Table 4.2.

Table 4.2 Teachers’ sets of assessment principles (credos)

CTA	CTB
<ul style="list-style-type: none">- “Sandwich”—always give positive, then negative/things needing improvement, then positive [A3:52]- Give opportunity for self-correction [A6:25]- Do not embarrass or humiliate with feedback [A4:18]- Distribute attention fairly... [A6:47]- but make sure to pay special attention to weaker students [A5:29]- Balance monitoring and giving space [A2:72]- Quantifying helps objectivity [in the context of summative assessments] [A1:41-42]- Give correction, because students like correction [A6:19]- Give feedback because it motivates (but doesn’t necessarily translate directly into language learning) [A6:25-33]- Do not demoralize students with too much negative feedback at once [A7:7]- Use [diagnostic] tests to get an idea of what students know and don’t know [A1:62-66]	<ul style="list-style-type: none">- Distribute attention fairly [B2:17]- Take time for feedback because it’s worth it [B6:45]- Affirm what the student did well (similar to “sandwich”, but this is from B asking student what was good and what should be improved, and students always say what was bad, so B always affirms first to counteract that; also, he doesn’t always end with a positive) [B7:7]- Balance monitoring and giving space [B11:15]- Do not demoralize students with too much negative feedback at once [B10:31]

As Table 4.2 shows, most of the principles related to either giving attention or giving feedback. This is in line with the teachers’ conceptions of assessment as being formatively focused. There is also a strong sense on both teachers’ parts that negative feedback could hurt students, and the possible damage needed to be compensated with praise and/or positive feedback.

¹⁰ I wish to thank my supervisor for this term.

One particularly interesting finding in CTA's data was a seeming conflict between a principle of feedback and a belief about language learning, as the following excerpt shows:

A: Well, I think the language they're learning in that lesson is probably completely different from the language I'm teaching them. What I'm saying, they don't even remember it; there's also things which I'm not even aware of now that they've probably learned. Probably from the organizational language of the classroom, from the metalanguage, but not from the actual teaching points. Because I don't think language acquisition works in the way that teachers think it works in the classroom.

M: But you give the feedback in the classroom because...

A: [pause] To keep the students happy, and because it's really an artificial construct. It's a conspiracy, really, between me and the students. They're happy with it and I'm happy with it, yes, it's a bit of a conspiracy. But I don't think it really... I think if you actually observe and research language acquisition, it doesn't work in that way. If you tested those students on those structures next week, it would probably *** them. I mean they'd probably acquire them but way down the line, or they'll never acquire them.

M: But you give the feedback partly because they want it...

A: Because it feels right. It just feels very good to give feedback. I'm contradicting myself now, but it just feels very good to give a lot of... I like to hand back marked work, and I like giving feedback on what they've said. The reason is because it means I've got active students, not passive ones; they have to produce something. But in fact from what I've read and think about it, a lot of language learning is receptive. They're probably learning as much language by listening to me as they are by speaking in class. ... I'd hate to think my classes are ever boring; I hope they're not, and it's by having this variety of input and output in the class that I try and maintain interest.

M: So just giving feedback partly because of their interest in it?

A: It keeps motivation and interest going, which in itself is a good thing, but I'm not sure if that's actually the language being learned. [A6:27-33]

Conflicts between beliefs and/or principles, often arising out of situational considerations, have been recognized in the teacher thinking literature (Clark & Peterson, 1986). In this case, I think the teacher is giving priority to the feedback-related principle over a belief about language learning, and does so because the former is actionable while the latter in a sense cannot be acted upon. That is, the teacher can impact the amount and quality of feedback (and thus, in the teacher's thinking, encourage student motivation, which helps learning) but can do little or nothing about what aspects of language students actually acquire; hence, giving feedback takes precedence.

II. Constructs applied in interactive assessment

I often asked the teachers about what they were looking at when observing students during pedagogical activities and formative assessment activities (e.g., individual speeches). Their responses could be classified as language-related, syllabus-related, and teacher-related aspects.

Language-related aspects, particularly pronunciation, vocabulary, and grammar, were predominant, as one would expect in a language class:

M: When you're looking at SC and SY, would you say you're listening in particular or are you just kind of listening in general to what's going on?

A: I'm looking for listening for pronunciation features, because they want to improve their pronunciation, and the language they use.

M: Anything else you're looking for?

A: Lexis, grammatical accuracy, strategies.

M: Strategies...

A: Conversational strategies, like re-formulation when they speak to each other if they don't understand each other.

M: Anything else?

A: That is all. [A2:51-59; also A2:68; A3:10-12]

B: [recalling observing a student speech] I'm tuning into the main features of his pronunciation and assessing to what degree they're interfering with the understanding of the *** message. And I make notes every time there's a word or a phrase that I can't hear clearly, or is a repetitive pattern; in his case the "z" "z" and "z"-so there's no distinction between "f" and "v" and "s" and "z". So I'm making notes, I make a few specific examples of it; so when I understand what seem to be the main problems of his pronunciation features, so that now I can sit down and give him feedback on that [B5:4; also B3:13; B6:39; B6:49; B7:5]

While the above were mainly in the context of speaking, the teachers also assessed listening (often through observing students' reactions to what the teacher or a classmate said [A2:22; B3:57]), reading [A1:84; B2:21] and writing [A3:72; B4:11].

The teachers also considered what can be called syllabus-related aspects of students' performance: teachers assessed students on skills that were itemized on the syllabus, like giving a presentation or summarizing a source (see Appendix 7).¹¹ The teachers

¹¹ Of course, several syllabus items such as "turn taking" and "recognising transition and signalling devices" can also be seen as linguistic in nature, depending on one's definition of the language construct.

looked at these aspects often in the context of an extended student performance (e.g., a speech [B3:13] or a group discussion [A7:118; B11:22]) or pedagogical activity where students were working individually or in small groups (e.g., a summary writing worksheet [B8:5]). This is understandable, considering that such contexts allow the teacher to observe a student's performance relatively at length.

By teacher-related aspects, I mean that a teacher may look for aspects of a performance that are not explicitly related to language or syllabus points, but rather to aspects that the teacher believes is important. CTB mentioned that he was not only trying to teach them the syllabus content, but also critical thinking, which he believed was important for his students in their future academic work [B6:27]. Thus, he also looked for this as he assessed and gave feedback to students [B6:29]. These comments were supported by the class observations; in several classes he set up activities that were ostensibly for the purpose of getting students to think critically about, for example, implicit assumptions or fallacious arguments in a text.

III. Stereotyping

Analysis of the data also seemed to show that, when making assessment judgments and comparisons, the teachers used what may be labeled as “stereotypes”, not in a negative sense but in the sense of having general preconceived ideas of what a student or group of students will be like, including linguistically, often based on experience with similar current and past learner groups. These stereotypes could be based mainly upon ethnicity, nationality, or language (e.g., Chinese speakers [B6:41], German speakers [A see below; B3:12-14]) or to a lesser degree upon course groups (e.g., English course for au pairs [A6:66], Foundation/Pathway course students [B4:9]); they also seemed to influence teachers' interpretations of a student's performance and to shape the feedback that was given. Two illustrations are provided below:

M: You mentioned you wanted to monitor the weaker students; was that something you'd resolved to do before the class?

A: Yes, I'd been thinking about it because they're the ones who are keeping quiet in the whole class discussions, group work's a chance to hear what they can do, how they can perform. And they're the ones who really should have the priority for the inessionals; the Germans don't really need inessionals, they could manage without it-many of the Germans.

M: Are they required to take it?

A: No, sometimes, very occasionally a department's said we want you to do an insessional, but they elect to do insessionals usually. ... The German ERASMUS students, they're over here on the exchange program, some of them are here more to advance their English than they are to advance their own subjects; English is more important because English is going to be vital to them in the future whether it's a future in commerce or whether it's an academic future. ... I'm managing my time [when paying attention to particular students during class]. Really, I try to be fair to every student in the class, ... to give my time and attention to everybody in the class equally, but perhaps a bit more for the ones who need it most, and they would be the Asian students. [A5:28-39]

M: You were able to pick up and diagnose some of his pronunciation errors. Where did that ability do you think come from?

B: In his case, I'm a pretty proficient French speaker, I lived with a French woman for 15 years, I've worked in France, I'm quite well-tuned in to the features of French accent. In the case of other students like Japanese students it's because I've taught a lot of Japanese students. In the case of an Iranian student like SS, I would not know what the primary L1 interference features would be, so there's two things I could do. I could listen to her and do the same process of making notes down and try to generalize if I've got any other Iranian students, and the other thing is to go downstairs and get a copy of Michael Swan's book called *Learner English*. And this book basically has a chapter by chapter account of the systematic interference errors for various L1 original languages. So it will give you the basic pronunciation problems you can expect to encounter with a typical say Chinese speaker, give you the typical grammatical structural problems. But what you should do is use it as a guide; you use it as a guide because you never assume that a student will have the typical features of their language. You should be listening to that student. So with SS and SI, if I'm going to listen to them, I'd have a quick look; if Farsi is included in that book—I'm not sure it is—I'll have a quick look at what interferences I can expect from a Farsi speaker. If I have any memory of working with Iranian students, then I'll try to remember that as well. [B5:7-8]

Actually, it is not clear from the data whether stereotypes are used only in interactive cognitions; CTB referred to them in interactive assessment, but CTA referred to them when discussing planning for interactive assessment (which is the topic of the excerpt from CTA above). This needs further investigation. However, it is clear that stereotypes do have important functions. This may be because stereotypes can reduce the teacher's cognitive load or mental burden—it would seem to be easier and more productive to rely on stereotypical knowledge than to start from nothing when assessing and giving feedback to a class full of students. It could also be because they are a kind of reference model (Orrell, 1995) or a form of typificatory scheme (Leung & Teasdale, 1997) and thus a natural human way of organizing data. I discuss some further implications of stereotyping in 4.8.3 below.

IV. Projection

The teachers occasionally described how, when observing a student, the teacher imagined that student in a hypothetical situation and made judgments based upon that. For example, CTA imagined what other native speakers in academic contexts outside the language classroom would think of the student's performance:

M: When you notice students' levels, is that according to some kind of...

A: Yeah. I'm thinking of how I would expect them to be performing in a seminar with other British students, from my experience with school education, where I've been doing my master's. I know what average British students perform; I'm assessing how well they would function in an academic seminar, how much listening stress there would be because of their accents, because of their inaccuracy, lack of coherence-cohesion...

M: As a person listening...

A: On a native speaker listening to them. [A2:61-64]

In a later session, there was a brief discussion about CTA's experiences on a master's degree course in TEFL, where English students sometimes avoided working with non-English students in discussions because of language issues:

A: I think I was writing there "do you mean 'are you saying?'" which I give in the feedback later. So I'm rating their performance against what I would expect, or against an image which I have of a real seminar, and I suppose I'm using my experience of seminars [on the master's degree] course, and my own undergraduate days—a long time ago now—but when I was a student.

M: You mean it comes to mind what you experienced in those days when you had seminars? You said there's an image of a ...

A: I think more recently, especially, mixed seminars with overseas students and home students together. I think if overseas students use appropriate language, then they're more convincing, it increases their credibility and they can interact with home students, because it's very very easy for overseas students to be excluded from seminars, from discussion, very easy. [A3:54-56]

CTB also expressed something like projecting how students would perform on an in-session summative assessment:

The others I was listening more with a view to what I described to them at the end, which is that at the end of this course, they will probably mix with another teacher's group, sit in groups of six, three from my group, three from the other teacher's group, and me and the other teacher who have a set band of descriptors, will probably watch a bit of video and then they'll be given a discussion rubric sheet, and then off they go; they're having their discussion in the middle of the group of me and the other teachers are *** so I was thinking about that with the others. While I was collecting input the strictly language based and pronunciation

impressions of the three that I hadn't heard before, I was also listening to the others for the other criteria that are in the assessment of discussion skills, which is courtesy towards other people, recognition of the response of the group, but mainly good presentation of arguments and ideas. That's what I was thinking about that, and generally looking forward to the assessment at the end. [B3:27; also in B3:53, displayed in III under 4.5.1]

Like stereotyping, this finding requires further study. But at this point, I would venture to guess that projecting students into a hypothetical situation may be a cognitive shortcut; perhaps it is a visual means of realizing standards of judgment. In other words, a teacher may visualize the student in an imagined situation and then judge his/her degree of "fit" to that situation. One possible advantage to this method is a kind of ecological validity; the teacher is familiar with a particular target language use situation (Bachman, 1990), and is aware of how a student may be judged in that situation.

V. Mental portraits of students

Naturally, as teachers gathered information about their students in a variety of classroom situations over time, they formed in their minds what can be called "mental portraits" of each student. Examples in the data can be found at [A5:25; A7:121; B2:17; B6:13; B11:11], among others. Both teachers described this stored information about particular students as "impressionistic" [A7:137; B2:21], most probably because the source of such information came not from systematic planned assessment but from incidental assessment within pedagogy, as the following excerpt demonstrates:

A: I've got SA, who I'd rate as very high level, and there's one more high level one there...

M: How did you gather that?

A: Oh, just hearing them speak and how quickly they can respond to questions in the English they use.

M: How's that?

A: In the English they use; the language she uses is a very advanced level. [A2:20-24]

These portraits served as an important basis for decision-making. For instance, the teachers compared members of a given class—who was relatively stronger or weaker than the others in different linguistic areas—presumably based on such portraits

[A3:39; A7:138; B2:6-10, 15-17; B8:13]. and such comparison was often done to group particular students or to decide whom to pay more attention to.

This finding resonates with what Harlen & James (1997) have argued in general education, that teachers build up tacit knowledge of students through a variety of means over time. There are important questions, however, about the quality of teachers' impressionistic knowledge; I discuss this below in 4.8.1.

VI. Assessments not related to language

Teachers obviously assess many things in the classroom, but my main concern in this study is with teachers' assessment of students' language use. Having said this, I have to recognize that when asked to share their thoughts during classroom practices that assessed language, the case teachers did speak a great deal about assessing non-language-related aspects of both individual students and the class in addition to evaluating student language use, particularly during observation of students doing pedagogical activities; among other things, they talked about assessing personalities and the dynamics of the class (e.g., how students work with each other, whether they are active or bored) [A3:37; B8:33], time constraints [A2:32; B5:18], their own teaching [A6:35; B4:21], syllabus content [A6:35; B4:13], and the appropriateness of the task [A5:25; B2:31].

It is very likely because of the vast and diverse concerns that teachers attend to in their observations of group work that they describe their mental portraits of students as impressionistic; the teachers seemed to mention the above concerns somewhat less and linguistic concerns somewhat more as the assessment practice was more focused on a particular individual or group for a longer period of time.

4.5.3 Uses of assessment information

As teachers gained knowledge of their students' linguistic abilities, they employed that knowledge for a variety of purposes. One of those purposes was for classroom management; the case study teachers often made decisions about how to group students at least partly based upon their knowledge of students' language levels:

A: [recalling his thinking during a group activity where students had to give a summary of a text] I picked on that group, I chose [SL] to listen to

first; well, she was the Text A reader, so she had to go first anyway. I was bound by the jigsaw nature of the activity. I had to have students who'd read the three different texts in the three different groups, and student A had to go first because there was a chronological order, there was a sequence to the texts. I could have chosen her or SE, because they both had the same texts. I chose her because I knew she was weaker from the work she had done here prior, at the early stage of the lesson. [A5:25]

B: I then have to make sure that they are talking, so when I'm sitting down here, usually, I mean in this case we had two groups of two, two pairs, one group of three, so I quickly had to make a snap decision as to what was the best situation: who should I put in pairs and who should I put in three. ... So basically the reason that I chose SY and SI from Iran and Japan to share the task of presenting one of the texts to SJ, is that these three students were noticeably in the first part of the lesson of a slightly lower level in communicative ability and in extracting information from the texts. [B5:4,6]

CTA also used what he knew of students to make judgments about who to attend to during class [A5:27-29], although the principle of fair distribution of attention balanced this (see Table 4.2).

Assessment information also informed lesson planning and the choice of activities for future class sessions; for example, CTB thought he could use more challenging and interesting materials and activities because his students were linguistically strong enough to handle them [B3:27-33; also A1:67-68; A5:40-43; B2:34-35,37-40].

Another use for assessment information was for giving feedback. This could happen immediately after assessment of student performance (usually during pair or group work):

B: [recalling his going around to each pair or group during a writing activity] I'm trying to look at [SA's and SC's] texts and get a very quick overview and say the most pertinent or useful thing and get them to do something with it, and in their case, I just asked them what do you think the differences are between your-I looked at it and I had a kind of first impression that I didn't think, that I just didn't feel on top of it to explaining it clearly, so I gave it back to them and I asked them what they thought the main differences were, and they confirmed what my impression was. ... And SA felt that SC had used a lot of detail, whereas she felt that he had been more general. And so that was basically my impression, so I said to them OK, and I said to SC that's what I felt about yours, and I wrote down the expression "arbitrary facts" for them. And so SC, yours seems to pick its way through, picking out arbitrary facts, but you've got a lot of general examples in there. So maybe we need summary statements there. And SA, you make your general statements but you don't tune them in with examples that will put them in the reader's mind. So can you now work together and try to produce that, and that got them going. [B8:17; also A4:18; A6:11]

or it could happen after some delay (like after the student was done giving a presentation) [e.g., A3:21; B2:35]; in these situations teachers often jotted down feedback to give to the student later. It was during these times that several of the assessment principles described earlier (see Table 4.2) came into play.

While the information teachers gained about students from interactive assessment was predominantly used to achieve pedagogical goals such as a positive and engaging classroom environment and student learning, the information also potentially played a role in summative assessment, at least according to CTB. The teachers' assessment information from the classroom were not directly incorporated into students' final course marks for the case study insessional classes. However, for at least one of the end-of-course assessments, the group discussion, there were usually two teachers who acted as raters, with one often being the class teacher for some of each group's participants. While this was relatively unproblematic in that both teachers were usually in close agreement on most ratings, teachers' knowledge of students became important when there seemed to be a discrepancy:

B: [talking about rating the summative group discussion assessment]
What usually happens then is if there are any so-called jagged profiles, if there are any students who don't fit easily and quickly with first impressions into a particular band, you usually have a quick discussion with the other teacher, and the teacher that knows them will be able to confirm or challenge your assumptions.

M: How often does that happen?

B: It's quite routine. Last time I did it with [teacher R], we did half of his and half of mine. One of the students was very interesting because he sat and said nothing throughout most of the discussion, so that I was unable to assess him any way at all, and then suddenly about three quarters of the way through he seized upon a point that somebody had made and then very rapidly demonstrated that he'd understood, that he had been listening to the whole discussion and he was then able to very eloquently present something that was wow yeah he's all there. You have to bear in mind therefore that some students who adopt a passive listening approach to discussion are actually engaged and contributing, and it may just happen that a student from another group that I've never listened to won't give me enough to put him in a band, so I'll ask his teacher probably in the room and I'll say I can't put him anywhere, he didn't say enough; what's he like when he talks? So, strictly speaking, we have to assess them on their performance in the given situation, but I think if somebody makes enough of a gesture in the direction of communicating and one of the teachers has an awareness of having seen him do this, I think it would be unrealistic to say it didn't affect the marking of it. [B3:55-57]

This excerpt shows how the knowledge teachers had of students based on interactive assessment could input into summative assessments, albeit indirectly.¹²

While not evidenced in these cases, teachers’ impressionistic knowledge of students also could be used for informing qualitative summative assessments of students; CTB said preessional courses he used to teach required teachers to write a brief qualitative description of the students’ language abilities, to be sent in the report to the students’ departments [B10:42-44].

4.5.4 *An integrated illustration of assessment practices, cognitions, and use*

In the preceding sections, I describe and discuss several thought patterns or processes that seemed to emerge from the data, especially in interactive assessment, and the uses to which information from such cognitions were put. It may be helpful to see an extended excerpt of data from the stimulated recalls, so as to illustrate both the nature of the data and how several of the above cognitions and uses are deployed in interactive assessment thinking for the purpose of feedback and classroom management (Table 4.3).

Table 4.3 An example of a teacher’s recall and accompanying cognitions and uses

Recall excerpt	Cognitions and uses
<p>[CTB is recalling a segment of class time in which he is observing groups of students working on an activity he has given them. Cognitions and uses are underlined.]</p> <p>B: <u>But basically they were all very on task, which is very good, it means it’s going well,</u> so in that case I will be sitting down, and here you can see me, I’m sitting between SC and SE, who are exchanging on one side of me, and SI, SY and SJ on the other. There’s two reasons for that. <u>One is from that position I can alternate between those two groups, and two is that SN and SS I’d made a fair degree of notes on their pronunciation and fluency in the previous class,</u> whereas I hadn’t on SC because she wasn’t there, and I hadn’t on the two Spanish, because they weren’t there. SY and SI I’d had, but it gave me a chance to tune in on the three that I hadn’t. That’s why I sat myself there. <u>Now what I’m basically doing is I can hear all of the groups, so I’m going from one to the other.</u> I’m tuning in to the two here, I’m tuning in to the two on my right, making individual notes about features of their pronunciation, their level of</p>	<p>assessment not related to language</p> <p>constructs applied; mental portraits</p> <p>assessment principle: distribute attention fairly</p>

¹² CTA said only a student’s performance on that summative assessment was allowed (e-mail, 21 June 2005); this difference could indicate differences in “local” application of organizational assessment policy.

Recall excerpt	Cognitions and uses
<p>fluency, and <u>for example SC I would be writing down that she speaks in a very slow and staccato fashion. Her language is quite accurate but she needs to improve her fluency. And like a Chinese speaker, she also pronounces every syllable equally when she speaks English.</u> That's what I wrote down about her. SE, it was then that <u>I noticed his level of fluency and his range of lexis are quite low relative to some of the others in the group,</u> and that's when I realized listening to SJ on this side that actually he was more eloquent in this context compared with the other Spanish guy than he had been when I'd been *** on his *** stuff.</p> <p>M: That's really interesting.</p> <p>B: So he then demonstrated quite a rapid fluency. I also wrote something down, quite a structural fluency in his argument. Now if we take a step back and think about generalizations that you might or might not have in your mind when you're making these snap decisions, if I step back, I'll realize that of course it fits a kind of stereotype of different language learners, that Asian language learners like the Chinese and-I haven't got a stereotype for SS and SI, but I can see SS in the same context as SC, but it wouldn't be that kind of stereotype-but certainly with Europeans, like Spanish and Italian students, they have a verbal, an oral confidence, a willingness to express themselves in spoken language, which makes them fairly fluent, with lots of errors, whereas Chinese education system, Japanese education system will tend to restrain--and maybe there are cultural factors as well-restrain their willingness to experiment and make mistakes, so that it's interesting that I noticed he was good when he was speaking freely, and noticed he had problems when he was analyzing text. So all of these things are spinning around in my mind, and I'm also walking around to the other group, but I don't feel a need to write down what they are doing orally—SS and SN—because I feel <u>I've got them in my mind fairly clear.</u> The reason I'm walking around to sit with them is because <u>fair distribution of teacher attention in the class is an important thing to do.</u> So even though I don't really need to take notes on them, I'll do it just to confirm my prejudices and I might pick something up that I haven't noticed. But I was mainly concerned with the two Spanish guys here and with SC to get good preliminary impressions of them, and then also there's a certain spontaneity because I will allow myself to join in the conversation if they're asking me an open question. If they're asking me an open question about language, I'll do my best in situ to target—as you probably saw me do when they were looking at the text—I'll target the language and I'll ask them what's it doing in the sentence. What's the function of this piece of language that you've got, and I'll try to make them guess it from context before I'll give it to them. But I will give it to them in order to facilitate the communication, there's no point blocking them for no reason. And also with these two as you can see me there I got quite involved with their discourse. Now even as I do that, I realize what I'm doing is I'm relaxing and I'm allowing myself to get into the conversation. What I'm doing is in a sense monopolizing</p>	<p>use: delayed feedback</p> <p>constructs applied: mental portraits stereotype</p> <p>comparison within class based on mental portraits</p> <p>mental portraits</p> <p>assessment principle</p>

Recall excerpt	Cognitions and uses
the conversation that they're having, which can be a good or a bad thing; really it's about them talking, but at the same time it's also about them getting feedback on how well they're communicating. And because they were engaging at quite a high level with this argument they were having about what the statistics represented, I was able to join in on that, as you can see me there, demonstrating and waving and getting very involved. And it's rewarding for them because they're engaged in a full conversation with a native speaker teacher. It's very on task, it's very related to the materials, and it's showing them that they're functioning at this level. <u>And with the others I didn't get engaged on that level so much because they were probably putting together the bits and pieces that they were trying to communicate to each other a little bit more slowly.</u> [B2:15-17]	use: immediate feedback comparison within class based on mental portraits

The above excerpt illustrates the cognitions and uses in action. While not every video segment of a classroom assessment practice elicited such variety, nor did every segment elicit an assessment of students, combination of such cognitions were often in play in most of the segments.

4.6 RQ4: What are the sources of these cognitions?

This question turned out to be very difficult to answer specifically. The main problem was conceptual, as I later realized: the above cognitions are categories I formulated based on analysis of the data, involving generalizing from specific instances. Hence, it was not possible to ask about the source of a cognition/category, since it was actually comprised of several data instances, each of which had a source. For example, it is not possible to assign one particular source to stereotypes; based on CTB's data, he had a stereotype of French learners based on his personal experience of working in France, etc., but he had a stereotype of Chinese learners based on teaching experience and reading.

Another problem was that some cognitions seemed so taken-for-granted that I did not ask about their origins; to do so would have seemed odd: "what are the origins of your considering classroom parameters?" or "where did you get the principle of 'distribute attention fairly'?"

Finally, while particular activities that involved assessment could be pinpointed relatively easily, the assessment cognitions deployed when executing the activity

may not have come from the same source, or may have involved several cognitions. For example, both teachers had students participate in group discussions during class time. The discussion activity was based on the summative group discussion activity that manager EPM put into place for the preessionals and that was extended to the insessionals [A7:82]. Thus, the activity source appeared to be from a colleague/manager. But the cognitions deployed—principles, constructs, stereotypes, projections—did not come from the same source.¹³

Hence, in providing the findings for RQ4, I can only discuss in a collective way the sources of assessment practices, strategic considerations, and interactive cognitions. Given these caveats, the data that was collected from the interviews and recalls pointed to a few prominent and key sources: language teacher training, teaching experience, books, colleagues, and other personal experience.

Language teacher training was perhaps the most important source, because it gave the teachers a particular methodology for language teaching (and, by implication, for assessing) and provided them with beliefs about language learning¹⁴:

B: [Setting up groups and then observing them] comes very strongly from the training that I've had, so that I feel better as a teacher when I'm doing that kind of thing than I do standing at the front. When I stand at the front, I think "I'm lecturing now."

M: What was it in the teacher training that...

B: It did all of that; I wasn't taught like that when I was a kid doing subjects; teaching was from the front. I realized a hell of a lot of research had gone into how people learn languages and designing the way ELT is supposed to be done: get them using the language, step back, get into a humble position and facilitate that. And I thought that worked really really well. [B4:23-25; also A4:5-6]

M: Where do you think you kind of picked up listening to students as they do a role play and giving them feedback?

A: Where did I pick that up? That was when I was on my diploma; no actually; yes, it was from presentation-practice-production, where you withdraw at the production stage *** and give the feedback later; you

¹³ It is possible that particular activities "afford," or enable, particular cognitions while foreclosing others (see Ellis, 2003, chapter 6, on sociocultural theory applied to language classroom tasks for background), and so learning a new activity or by extension a new practice would involve learning new cognitions; however, such cognitions might not be new in the sense of created from nothing but new in the sense of adapting previous cognitions to the new activity/practice. This topic is not explored in this study.

¹⁴ It could even contribute to projections, as CTA's data suggests.

don't interrupt in the production stage. So that was when I was doing my RSA diploma back in '91.

M: ... When you say that was during your diploma training, did they give you examples, or how did that work? Or was it through teacher training?

A: It was just reading the books on the reading list of the course, and... I was observed twice by my tutor... It's mainly through reading. There's a book called *At the Chalk Face* ... which gives lesson plans. I got it from that book; it was on the reading list for my diploma. Not through observing other people; I wasn't observing other people at the time. [A6:12-15]

As alluded to in the above data, books were also a significant source, both during training and during teaching experience. They contributed to practices [B6:25] and to cognitions such as stereotypes [B5:8].

Teaching experience provided a wealth of information, especially about tasks [A6:68]—for example, the case teachers often experimented with new activities, including assessment tasks [A1:70-76; B4:21], and their experiences with these influenced whether and how they would use them in the future—and about stereotypes. One could say that teaching experience provided baseline knowledge (about what new students would probably be like, what tasks would likely be successful, etc.) that could serve as a basis for assessment judgments and that could in turn be amended based on new experience.

Colleagues were mainly a source of pedagogical activities, but they were also a source for CTA when designing the summative essay exam; it developed from “discussion with colleagues, ... and you'll be looking at what others have done at other institutions. Also, speaking to, meeting people at conferences and professional interest meetings and discussing how they assess.” [A7:24]

Finally, personal experience also clearly influenced teachers' assessment thinking. In CTA's case, it shaped his views about assessment:

M: What do you think has kind of influenced your thinking about assessment? ***

A: My own experience as a learner.

M: Can you give me some specific examples, does anything come to mind?

A: Well, most recently, my master's, getting feedback on my assignments, their being marked and being rated. Being rated on practical ways, on a

sports course—skiing, rock climbing. ... Yeah, it's nice to get a 3-star silver on my beginner's skiing, and to be told that I was fair to middling. It's quite interesting to listen to the way sports instructors, sports coaches, assess people. I was having a discussion about this with a fellow EFL teacher who's also a rugby referee and referee trainer, and we're trying to decide *** what a fair is, how much of a good, or what's the difference between a poor and a fair. Because we have to... there are similarities especially when you're assessing students' oral skills, it's a bit like watching someone play a sport or climb a mountain, climb a rock. [A1:16-21]

In CTB's case, his “stereotype” information about French learners of English was based on his being a fluent speaker of French, having worked in France, and having lived with a French woman for a number of years. [B5:8]

A couple of important implications can be drawn from these findings on the origins of assessment cognitions. First, language teacher training can have a key role in teachers' future assessment thinking, mainly through teaching methodology but also through providing resources for assessment practices and activities. Although the effectiveness of initial teacher training is debated in the literature (Borg, 2003), the data from Stage 1 shows how potentially significant an impact it can have when it comes to assessment thinking. If one wants to improve classroom assessment, then teacher training is clearly a leverage point; however, the above analysis suggests that teacher trainers need to give thought to how particular teaching methodologies align with particular assessment practices, so that student teachers will not find one or the other irrelevant.

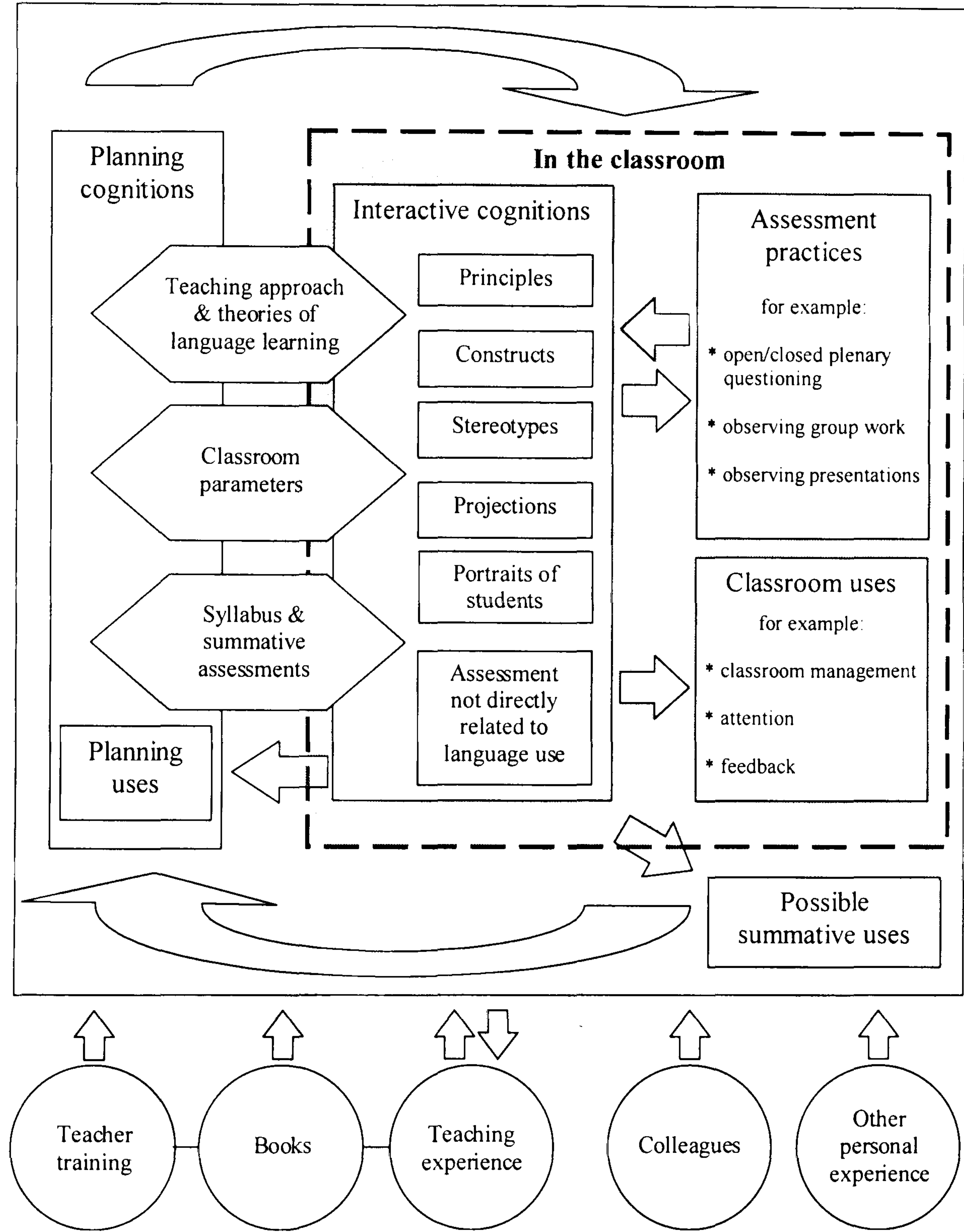
Second, it could be argued that teachers may become better assessors when they have increased “stereotype” information from teaching experience, books/training, colleagues, and personal experiences dealing with language learning and different cultures. These could allow a teacher to have a broader and/or deeper knowledge base on which to build judgments and inferences and from which to draw feedback information.

4.7 General discussion and summary of findings

In summary, the Stage 1 data revealed that while language teachers have a number of assessment practices in common, the frequency and nature/quality of those practices

are unique to individual teachers teaching particular classes; this is due largely to a variety of assessment-related cognitions engaged in by the teachers. Figure 4.2 displays these cognitions and the relationships between them, in what I have termed a teacher’s “assessment cognition network.”¹⁵

Figure 4.2 Proposed model of an assessment cognition network



¹⁵ It should be kept in mind that, while I refer to it as an entity, it is in fact only a heuristic model; I think it is more than a mere personal construction (which an extreme constructivist paradigm would suggest), but it is still a construction, albeit one that attempts to represent in a limited way actual cognitions and their relationships in teachers’ minds.

Represented by the hexagons in the figure, overarching considerations regarding personal teaching approach and views on language learning, syllabus and summative assessment requirements, and other classroom parameters (like number of students) are fundamental (see 4.5.1); they impact thinking both in lesson planning and during class time. Also as the figure shows, thinking during planning and class time are different but related; the former sets up much of the context of thinking in the latter, while “data” from the latter afterwards contributes to thinking in the former.

The figure also indicates the cognitions employed during interactive assessment: principles, constructs, stereotypes, projections, mental portraits of students, and other assessment not directly related to language use (see 4.5.2). These cognitions and the considerations mentioned earlier underlie the assessment practices that can be observed in the classroom (see 4.4); information gained from such practices are put to particular uses, especially in the classroom (see 4.5.3) and may also lead to development of those cognitions and considerations (such as portraits of students or stereotypes).

The circles at the bottom of Figure 4.2 represent the sources of assessment practices and cognitions: teacher training, books (accessed both in the context of teacher training and teaching experience), teaching experience, colleagues, and experiences from outside teaching (such as having lived in another country) (see 4.6). It should be noted that the cognitions in the large rectangle not only draw from but also contribute to—or constitute part of—teaching experience.

Three important points should be made regarding this model. First, the model can be said to overemphasize interactive assessment, which mainly involves observation of especially spoken performance, and underemphasize other common types of assessment, particularly homework and paper-and-pencil tests. This is mainly due to data collection limits; there were no stimulated recalls of assessment outside of class, and in these two cases the teachers did not give graded paper-and-pencil tests during the term. However, I would venture to guess that the cognitions in Figure 4.2 will also be applicable to teachers’ assessment work outside the classroom, too. For example, I presume the strategic cognitions of pedagogical approach, classroom parameters, and syllabus/summative assessment requirements would play a key role

in determining the frequency and nature of homework (seen as an assessment practice). In addition, it may very well be that teachers draw upon cognitions similar to those for interactive assessment when they assess outside class time. Davison (2004) asked her teacher participants to verbalize aloud what they were thinking as they assessed some essays, and then had them discuss in groups their assessment processes and judgments; in the excerpts of published data below, there are hints of the interactive cognitions:

Stereotype

A teacher reporting thoughts while assessing an essay:

“That should be 71, a safe C. Now, I don’t know. The language because though she used a lot of idioms, she does have a mass group of the language there. If it is overused, I think it is a very general thing with Hong Kong students. They swallow dictionaries and then they try and pump out as many of these sort of like ‘Every cloud has a silver lining’ and all of these sorts of stuff. I think it’s better, much, much better than any of the others that we have looked at so far. ...” (p. 321)

Stereotype and implicit construct

A teacher reporting thoughts while assessing an essay:

“For my students, I guess they can master the basics, at least the fifth level. I think they lack practice on logical thinking and organization of ideas, that sort of thing. So, I have a very heavy emphasis on organization and logical flow of ideas.” (p. 320)

Projection

A teacher R in a group discussion commenting on his/her own thinking:

“I have a dreadful conflict within myself as to what I call my intuitive judgment, which is what you are going on [when assessing an essay], I think. ... And thinking about the future and where is this child going, can this child cope? But that’s not what we’re being asked to do. We’re being asked to tick the box.” (p. 317)

This last comment from the Davison study leads to another point regarding my proposed model of teacher thinking in assessment, which is that the use of explicit criteria, frameworks, or bandscales may affect cognitions significantly. The Davison data suggest this; teachers who used criteria to make judgments seemed to verbalize a great deal of interaction with and interpretation of the explicit criteria, and did not voice the above cognitions, or saw them as somewhat opposed (as teacher R does in the above excerpt). This suggests that such frameworks do not make thinking explicit but actually lead to qualitatively different thinking. This is also hinted at in CTB’s data (see III under 4.5.1).

The final point about Figure 4.2 is that there are many elements that can be found in models developed from previous studies (see 2.4.1 and 2.4.5). For example, my proposed model and the model that McMillan & Nash (2000) propose for teacher decision-making in assessment seem to agree in the fundamental considerations of teaching approach, classroom realities, and external factors. In chapter 7, I consider at length the ways in which this study's findings, from both stages, relate to findings from previous research; suffice to say at this point that this kind of agreement provides a degree of external validity and of analytical generalization (Yin, 2003) to other educational settings. It remains to be seen, however, as to how robust the findings from Stage 1 regarding interactive cognitions are.

4.8 Issues and research questions for further research

Being the first phase of a progressively focused study, the data from Stage 1 was exploratory and broad. In the next stage, I planned to examine more deeply three substantial issues that were raised by this exploration; those issues and the research questions that were developed from them are described below.

4.8.1 Quality of teachers' impressionistic knowledge about students

Foremost is the issue of the quality of teachers' impressionistic knowledge about students. My interest in this stems from two concerns. First, throughout the data collection, the case teachers made comments about the language abilities of students in their classes. As mentioned earlier regarding mental portraits of students (see V under 4.5.2), these comments ranged from general statements (e.g., "X's quite good", "Y's weak") to specific strengths and weaknesses (e.g., "P has a strong vocabulary", "Q handles group discussion well"). However, I began to wonder how the teachers could know if their judgments were accurate; the case teachers had not mentioned this issue in the data collection. Second, in my reading of the literature on classroom assessment, it was apparent that there was disagreement about the quality of teachers' impressionistic knowledge of students. Some writers have argued that such knowledge is undependable and too easily biased; for example, Popham (2002, p. 50-51) states:

Although teachers are often forced to make inferences about students' knowledge, skills, or attitudes on the basis of informal observations, such

unsystematic observations sometimes lead teachers to make invalid inferences about a particular student's status. I'm not knocking teachers' informal observational skills, mind you, for I certainly relied on my own informal observations when I was in the classroom. Frequently, however, I was off the mark! More than once, I saw what I wanted to see by inferring that my students possessed knowledge and skills they really didn't have. Later, when students tackled a midterm or final exam, I discovered that the conclusions I had drawn from my observation-based judgments were far too generous.

Other writers have argued that such teacher knowledge can be very dependable and trustworthy, the main arguments being 1) that teachers see students perform in a variety of contexts, thus giving a more complete view of student ability, and 2) that teachers see students over a lengthy period of time, so any errors in judgment can be corrected over the duration of a class term (Harlen & James, 1997).

Thus one issue that will be explored in Stage 2 is the quality of teachers' knowledge of students; rather than attempting to judge that quality using some kind of external criteria (such as from psychometric theory—see Teasdale & Leung, 2000, about the contradictions that can arise when this is done; that is why I have studiously avoided framing this directly in terms of “validity” and “reliability”), I plan to examine how teachers themselves think their own judgments can be improved. The research question developed from this issue was:

RQ5. *How do English language teachers think they can increase the quality of their impressionistic knowledge of a student's language abilities?*

4.8.2 *The impact of managerialism*

The next issue I plan to follow up on in Stage 2 is the impact of increasing managerialism on assessment thinking. Again, my interest in this issue arises from a combination of empirical and theoretical concerns. I found that the two case teachers differed widely in their attitude toward the managerial changes; CTA expressed support for them, not only because he was the one who had produced the syllabus but also because he saw the changes as making the class more systematic and uniform than before [A7:44]. CTB, on the other hand, took a more questioning stance, at least towards the more explicit syllabus; he expressed feeling a great deal of time pressure from it in his interactive thought processes.

I suggest that these diverse responses to managerial action can be understood in light of their diverse pedagogical approaches. I wish to discuss this at the “micro” and “macro” levels, with the help of two analytic tools from the research literature. First, at what I see as a “micro” level, Torrance & Pryor (1998) propose two tendencies for assessment: convergent and divergent. Their characteristics are displayed in Table 4.4:

Table 4.4 Characteristics of divergent and convergent assessment (from Torrance & Pryor, 1998, p. 153)

<i>Convergent Assessment</i>	<i>Divergent Assessment</i>
Assessment which aims to discover <i>whether</i> the learner knows, understands or can do a predetermined thing. This is characterized by: <i>Practical implications</i> (a) precise planning and an intention to stick to it; (b) tick lists and can-do statements; (c) an analysis of the interaction of the learner and the curriculum from the point of view of the curriculum; (d) closed or pseudo-open questioning and tasks: CTA (e) a focus on contrasting errors with correct responses; (f) judgemental or quantitative evaluation: CTA (g) involvement of the pupil as recipient of assessments. <i>Theoretical implications</i> (h) a behaviourist view of learning; (i) an intention to teach or assess the next predetermined thing in a linear progression; (k) a view of assessment as accomplished by the teacher. This view of assessment might be seen less as formative assessment, than as repeated summative assessment or continuous assessment.	Assessment which aims to discover <i>what</i> the learner knows, understands or can do. This is characterized by: <i>Practical implications</i> (a) flexible planning or complex planning which incorporates alternatives; CTB (b) open forms of recording (narrative, quotations etc.); (c) an analysis of the interaction of the learner and the curriculum from the point of view both of the learner and of the curriculum; (d) open questioning and tasks: CTB (e) a focus on miscues – aspects of learners’ work which yield insights into their current understanding – and on prompting metacognition; (f) descriptive rather than purely judgemental evaluation; (g) involvement of the pupil as initiator of assessments as well as recipient. <i>Theoretical implications</i> (h) a constructivist view of learning; CTB (j) an intention to teach in the zone of proximal development; (k) a view of assessment as accomplished jointly by the teacher and the pupil. This view of assessment could be said to attend more closely to contemporary theories of learning and accept the complexity of formative assessment.

As the table shows, convergent assessment is characterized by such things as precise planning, tick list recording methods, closed or pseudo-open questioning, and an error-correction emphasis. Meanwhile, divergent assessment is characterized by, for example, flexible planning, open recording methods, open questioning, and an emphasis on student understandings and metacognition. Based on the practices and beliefs described earlier (see 4.4 and 4.5.1 especially), CTA and CTB can be seen as having somewhat convergent and divergent assessment tendencies, respectively, as the table also shows. This may explain their differing attitudes towards the managerial actions; a more explicit syllabus would support precise planning and a linear progression while militating somewhat against a more flexible approach, and time pressure would encourage more direct teacher control (to keep class time moving along and to cover the syllabus material) and discourage (time-consuming but potentially very educative) tangents and digressions characteristic of less direct teacher control. Furthermore, taking into consideration Figure 4.1, it may be that a more explicit syllabus also encourages more public forms of assessment rather than private ones, again because of the need to control time.

I would like to now set this “micro” analysis in a larger “macro” analysis: Broadfoot & Pollard (2000), basing their work on Bernstein, set out two contrasting models of education, described in Table 4.5.

Table 4.5 Contrasting aspects of competence and performance models of education (from Broadfoot & Pollard, 2000, p. 20-21)

<i>A 'competence model'</i> LIBERAL PROGRESSIVE EDUCATION		<i>A 'performance model'</i> PERFORMANCE EDUCATION	
<i>Schools and teachers</i>	<i>'Invisible management' with relative professional autonomy</i>	<i>'Visible management' with relative professional regulation</i>	<i>Classroom and pupils</i>
Organizational form	Professional, with flat manage- ment structure. Control through self-regulation, socialization and internalization of norms.	Mechanistic, with hierarchical structure and bureaucracy. Standardization for control and co-ordination.	Autonomy Limited.
Management style	Collegiate, with emphasis on proficiency, dialogue and consensus. Informality in relationships.	Managerial, with emphasis on efficiency and target setting for results. Greater formality in relationships.	Space Flexible boundaries and use. Time Flexible emphasis on present experiences.
Teacher roles	Teachers as facilitators, with affective dimensions seen as intrinsic to the teaching role.	Teachers as instructors and evaluators, with emphasis on cognitive and managerial skills.	Activity Strong control over selection of knowledge and explicit promotion of specialized subjects and skills.
Teacher professionalism	Professional covenant based on trust, and commitment to education as a form of personal development. Confidence and sense of fulfilment and spontaneity in teaching.	Professionalism is the fulfilment of a contract to deliver education, which is seen as a commodity for individuals and a national necessity for economic growth. Teacher confidence and fulfilment are less.	Evaluation Emphasis on immediate, present qualities using implicit and diffuse criteria. Control Relatively 'invisible', with control inhering in interpersonal communications and relationships.
Teacher accountability	Personal and 'moral' accountability.	External and contractual accountability.	Pupil products Pupil products are simply taken to indicate performance, as objectified by grades. Teachers instruct and assess using established procedures and criteria.
Whole school co-ordination	Relative autonomy and informal teacher collaboration.	Formal school planning with 'contrived' collegiality.	
Economic costs	Expensive, because of sophisticated teacher education and time-consuming school practices.	Cheaper, because of more explicit teacher training and systematized school practices.	Pupil learning Highlighting intrinsic motivation and encouraging mastery orientation. Potential for 'deep learning', but tendency to produce routinization and evasion.

As the table shows, a competence model is founded on professional autonomy, more internal controls, less formality, and considerable flexibility in the classroom, while a performance model has greater regulation, more external controls, more formality, and more restrictions upon classroom practice. While I do not have enough data to “place” on every aspect the language center in which the Stage 1 cases occurred, there are a few aspects in which a shift from a competence to a performance model seemed to be supported by the data I did collect. These are: from a relatively flat to a more hierarchical management structure, and from a situation where teachers had the autonomy and flexibility to teach and assess whatever and however they wished to a situation where the syllabus and summative assessments are relatively standardized, and hence a) there is relatively stronger structuring, sequencing, and pacing of class time due to the syllabus, and b) there is increasing use of established procedures and criteria to assess pupil products (the summative assessments).

My tentative argument here is that while a competence model of education allows for both convergent and divergent assessment tendencies, an education system with a performance model of education lends itself to supporting convergent assessment and discouraging divergent assessment. Therefore, in the example of this language center, since CTA’s overall teaching and assessment approach is more explicitly teacher-structured, he may feel less pressure than CTB, whose approach requires more flexibility.

This is mostly conjecture at this point. More generally, there may be what I term a “micro-washback” effect from summative assessments determined at the organizational (as opposed to the individual teacher) level. Thus, in the next stage of data collection, I want to explore this possibility further. The research question for Stage 2 based on this issue was:

RQ6. *How do teachers who have experienced managerial changes like a more explicit syllabus and more explicit summative assessments feel such changes have affected their assessment practices?*

4.8.3 *The use of stereotypes and projection*

The last substantial issue I want to follow up on in the next stage is the confirmation of teachers' use of stereotypes and projections in assessment (see III and IV under 4.5.2). As far as I am aware, there has not been coverage of either of these in the assessment literature. Perhaps the closest reference to stereotyping that I have found has been Becker (1952) in Broadfoot (1996, p.5); this study found that teachers categorized students in terms of stereotypes such as "lazy", "dull," or "bright," in order to provide "appropriate" teaching. However, these stereotypes seem to be broad attributions often about character, and each has certain positive or negative connotations; however, the stereotypes found in Stage 1 are more specific in nature, often regarding linguistic or cultural background, and are less connotatively "loaded" than the ones Becker found. In regards to projection, perhaps the nearest material I have found is in the teacher thinking literature, regarding the use of images by teachers to describe their work (Clandinin, 1986), although these again seem to be broad in nature compared to the projections focused on the target language use situation that Stage 1's teachers seemed to express. If teachers do use stereotypes and images in assessment, then this raises some intriguing questions, such as whether teachers' stereotypes and images serve as a kind of construct, or perhaps criterion-referencing, or whether the assessment skills of teachers in training can be improved through the use of images. Stereotypes and projections are investigated more deeply in Stage 2 through the following research question:

RQ7. *Do teachers use stereotypes and projection when they conduct interactive assessment of students?*

4.9 Summary

This chapter has covered the findings from the first stage of data collection and analysis, which involved case studies of two teachers of an EAP language class. The background to the cases was given, followed by a summary of the teachers' definitions of assessment. This was followed by a description and analysis of the assessment practices and underlying thought processes of the two teachers. It was found that the two teachers had differing patterns of assessment practices, most

likely resulting largely from their differing teaching approaches and beliefs about teaching and learning. It was also found that organizational policies in the form of a set course syllabus and uniform summative assessments, plus considerations of classroom parameters such as class size, played very significant roles in the teachers' assessment thinking. Several other cognitions were found to be used during interactive assessment; these included guiding principles of assessment, constructs, stereotypes, projections, and mental portraits of students. The uses to which assessment information was put were also outlined. These findings were followed by a discussion of the sources of these assessment cognitions.

Combining the aforementioned assessment practices, cognitions, uses, and sources, a model of an assessment cognition network was proposed. Finally, three issues for further research were identified: the quality of teachers' knowledge of students, the impact of managerial activity on teachers' assessment cognitions and practices, and the existence of stereotypes and projections as assessment cognitions.

The next chapter explains the methodology of Stage 2, which was designed to investigate these three issues.

Chapter 5 Methodology: Stage 2

5.1 Introduction

Following the Stage 1 methodology and findings detailed in chapters 3 and 4, this chapter presents the methodology of Stage 2. Similar to the order of chapter 3, I first provide the rationale for how the Stage 2 data collection was designed and executed and then elaborate on how the data was analyzed. As a reminder of how the two stages of this study fit together, I have repeated Table 3.1 as Table 5.1 below.

Table 5.1 Overview of research design (same as Table 3.1)

Stage 1	
Research questions:	
RQ1. What are teachers' definitions of assessment?	
RQ2. What are the classroom assessment practices of English language teachers in the EAP context?	
RQ3. What cognitions underlie these practices?	
RQ4. What are the sources of these cognitions (e.g., initial teacher training)?	
Strategy: Case study	
Data collection methods:	
* Classroom observation (RQ2) * Interviewing (RQ1) * Stimulated recall (RQ3, RQ4)	
Sample: 2 teachers teaching an insessional EAP course for 1 term in a university language center	
Timeline:	
Data collection	Jan 2004-April 2004
Analysis (including in field & writing)	Feb 2004-Dec 2004
Stage 2	
Research questions:	
RQ5. How do English language teachers think they can increase the quality of their impressionistic knowledge of a student's language abilities?	
RQ6. How do teachers who have experienced managerial changes like a more explicit syllabus and more explicit summative assessments feel such changes have affected their assessment practices?	
RQ7. Do teachers use stereotypes and projection when they conduct interactive assessment of students?	
Strategy: Interview	
Data collection method: Focus group interviewing (RQ5; RQ6, RQ7)	
Pilot: 3 doctoral students in TESOL with EAP experience in EFL contexts	
Sample: 3 focus groups, each consisting of teachers from a university EAP center or program	
Group 1: 4 participants; Group 2: 3 participants; Group 3: 5 participants	
Timeline:	
Data collection	March 2005-June 2005
Analysis (inc. in field & writing)	March 2005-June 2005

5.2 Rationale for design and conduct of Stage 2 data collection

5.2.1 Stage 2 research questions

As mentioned in 4.8, the Stage 1 analysis raised some further questions requiring exploration. Those issues and questions are reiterated below.

Issue 1: the quality of teachers' impressionistic knowledge of students

RQ5. *How do English language teachers think they can increase the quality of their impressionistic knowledge of a student's language abilities?*

Issue 2: the effect of increasing managerialism on assessment thinking

RQ6. *How do teachers who have experienced managerial changes like a more explicit syllabus and more explicit summative assessments feel such changes have affected their assessment practices?*

Issue 3: the use of stereotypes and projection during interactive assessment

RQ7. *Do teachers use stereotypes and projection when they conduct interactive assessment of students?*

Overall, it can be said that the purpose for Stage 2 was both to confirm or “triangulate” (see 3.5.3) findings from Stage 1 (as with RQ6 and RQ7) and to explore new issues hinted at in Stage 1 (as with RQ5). The following sections present how the preceding research questions were investigated.

5.2.2 Deciding upon an interviewing strategy

What Table 5.1 shows is a straightforward progression from the first to the second stage of my study; what it does not show is the rather arduous process by which Stage 2 came to be designed, for the Stage 2 strategy given in Table 5.1 is not the original design I had in mind.

I had initially planned to continue with more case studies outside the EAP context: following the replicative logic of multiple case studies (Miles & Huberman, 1994; Yin, 2003), I thought of looking at instructors teaching classes in adult ESOL and

private language schools, to see how changes in context would strengthen or disconfirm my Stage 1 findings. However, several reasons led me to change designs.

The first was that I simply could not gain access. I contacted at least 12 people, representing or having access to teachers of adult ESOL programs (in colleges or community education centers) and private language schools in at least four geographic areas; for one reason or another, no teachers who were interested and available to participate were forthcoming.¹ After several months of being unable to find participants, I thought a different strategy altogether would be needed.

One by-product of this difficulty was that, as accessing adult ESOL or private language school contexts became less likely, I came to realize after discussion with my supervisor that concentrating on the EAP context would give greater coherence and unity to the study and its findings. Also, as I mentioned in the introductory chapter (see 1.3), I was familiar with EAP because of my previous teaching experience in the US and Taiwan. Finally, while I still think the other contexts of adult ESOL and private language schools are worthy of investigation and in fact

¹ The experience of looking for participants was educational in itself. Some were contacted by “cold-calling” (i.e., having had no previous contact)—I knew this would be a low-success method, but it was the only option considering my lack of connections with particular educational institutions—while others were friends or acquaintances of colleagues. I told each contact (or left a message for them) that I was looking for research participants and (unless I was turned down immediately or not called back) gave a brief summary of what would be required, particularly in terms of time. In the end, the search was fruitless. A couple contacts did not reply. Some contacts tried asking people they knew but there was no interest expressed. Other contacts gave various reasons that prevented me from finding participants in their programs. One was organizational constraints; one contact said her institution had no policy in place regarding external observers, another said the school policy was that external observers had to pay a 25 pound fee per day, not including interview or recall time, while another said her institution was undergoing inspection and thus could only allow classroom observation without time for interview or recall. Another reason given was sensitivity to the fact that many students in adult ESOL classes were Muslim women and hence would not want to be observed by a male. But perhaps the most common response was that the teachers—many of whom were part-time—did not have the time to participate. As the methodology literature points out, inability to participate also signifies something; I think this situation among adult ESOL teachers reflects time and resource pressures that could have a significant impact on classroom assessment, although it is a pity this cannot be examined empirically. Ironically, one adult ESOL contact said her staff would have a lot to talk about regarding assessment because a new national policy setting up a qualifications framework for adult ESOL students, *Skills for Life*, was currently being implemented, but making all the changes left her staff with no time to participate in my research. After several months of being unable to find participants, I thought a different strategy altogether would be needed. In what turned out to be prophetic, one adult ESOL contact that I talked with early on said he honestly thought I needed to overhaul my approach, because no one would be able to spare the time required; perhaps I should have listened to him from the outset!

have been greatly underrepresented in the research literature, I had to keep in mind the limited scope and resources of this study.

The second main reason why I changed designs was that the nature of the questions I wanted to answer required a different approach from case study. While the replication logic of multiple case studies offered a rationale and coherence that appealed to me, I also took note of a warning from Yin: as the research questions develop during a case study, it may be that a case study methodology becomes less suited to answering those emerging questions. I noticed that my new set of more specific research questions required teachers a) to recall past events in their professional experience (e.g., for RQ6, managerial changes to a course the teacher had been teaching²), and b) to make generalizations about their thinking and practices (e.g., for RQ5, principles for strengthening the quality of teachers' impressionistic knowledge of students). In light of these requirements—and the need to keep participant time commitments minimal but meaningful—I thought a case study strategy would be inefficient, in terms of time and amount of data, or unable to provide what I needed, in terms of nature of data collected. I then decided upon an interviewing strategy.

As mentioned previously, interviewing can refer both to a general strategy and to many specific techniques. Its great utility stems from the fact that it provides an efficient means of accessing participants' attitudes, perceptions, meanings, and, in particular, personal experiences (Fontana & Frey, 2000; Punch, 1998). This aligned well with the kinds of data I needed to answer the research questions. In a sense, interviewing opened up to investigation a whole range of “cases”—in the form of teachers' memories of courses past and present that they had taught—that observational case study, being restricted to a small number of contemporary instances, could not access.

² Ideally, I would have liked to have used case studies of instructors teaching classes that were currently going through managerial changes (as the Stage 1 cases were), as this would permit classroom observation (see Wall & Alderson, 1996, on the importance of this for investigating washback). However, the difficulty of finding participants to make such a large commitment precluded this.

5.2.3 *Data collection method*

Given an interviewing strategy, I needed to decide how to realize that strategy. Specific interviewing methods vary along several dimensions (Bogdan & Biklen, 1998; Fontana & Frey, 2000; Punch, 1998); perhaps the most often mentioned are the degree to which the interaction and questioning is structured (ranging from highly structured and controlled interviews in, for example, surveys to loosely structured interviews in oral history or ethnographic research) and the number of interviewees involved (individual or group). Choosing the type of interview design to employ, therefore, depends on the research purposes and questions (Punch 1998). Because I had a fairly specific set of research questions, the choice of a semi-structured interview using questions derived from my research questions was obvious. However, whether to interview individuals or groups was less easy to decide. I explain my final decision below.

The Semi-structured Focus Group

Although there are several types of group interviews (Fontana & Frey, 2000, p. 653) provide a typology), it was focus groups—where a handful of participants discuss face-to-face their thoughts about a particular topic given to them by a moderator—that I was most familiar with; in fact, forms of focus groups can even be found among the studies reviewed in Chapter 2 (Davison, 2004; McCallum et al., 1993). Thus, it was between these and individual interviews that I was deciding. As I read more of the methodology literature, it became clear that focus groups were well-suited for my design needs. First, they could be used to generate feedback on previous research findings, “by returning to key stakeholders who are actually working with, or are invested in, the topic of our research and asking them to provide their interpretation of the findings or suggestions for further research” (Vaughn, Schumm, & Sinagub, 1996, p. 29); this was exactly one of the things I wished to do in Stage 2. Second, in comparison to individual interviews, “focus groups may have an advantage for topics that are habit-ridden or not thought out in detail.” (Morgan, 1997) This struck me because my research questions dealt with topics that were part of the everyday routine of teachers; it seemed to me that a teacher might not have much to say about such topics in a one-to-one interview, but another participant’s

comments in a focus group discussion could stimulate that teacher's thinking and provoke a response.³

Having decided to use focus groups and knowing that they would be relatively structured by my research interests, I then had to prepare and pilot the interviewing guide containing a brief questionnaire (for background information on the participants) and the questions I was going to ask (Morgan, 1997).

5.2.4 Piloting the focus group technique

The interviewing guide I eventually created is shown in Appendix 9. As it shows, I "operationalized" each research question with a set of interview questions.

I then contacted several of my doctoral classmates whom I knew had had experience teaching EAP. Three of them were able to participate in my pilot (the others were unavailable because of time commitments or illness); these three were non-native English speaker teachers who had taught EAP in further or higher education overseas. I was somewhat concerned because they only represented EFL contexts, not the EAL/ESL context of my prospective focus groups, but I proceeded because a) my main purpose was to see if my questions made sense and if the focus group interaction would produce the kind of data I wanted, and b) their perspectives could provide valuable insights, especially for further research.

The pilot focus group was held in a meeting/seminar room at our university and lasted for about 80 minutes. In the event, I was able to practice moderating the focus group with some key principles in mind that I would later use in future focus groups:

1) Stay focused on the guide questions but "go with the flow" too. As the focus group literature recommends, it is important to keep the discussion focused on the questions of interest, but the order or exact manner in which they are asked should be flexible and respond to the interaction as it proceeds, and follow-up questions are equally important to explore or clarify responses.

³ As Morgan (1997; 2002) points out, however, there has been little empirical work done comparing individual interview data with focus group data to support or disconfirm such a belief.

2) Give participants space to talk but give everyone an opportunity. One potential problem with focus groups is that one participant may dominate the discussion and thus prejudice the data. To avoid this, I tried to give each person enough time to talk and yet be firm in moving the discussion on to other participants and/or other questions.

3) Encourage participants to focus on specific experiences. Krueger (1994, p. 66). writes that getting them to “think back” and “time shift” can encourage more accurate responses, as participants center on what they have done as opposed to what they plan or wish to do. To do this, I often asked participants to give specific examples from their professional experience. Also, for some of my “set” questions, I handed out excerpts of data from Stage 1 (see Appendix 9) to make my own questions contextualized.

4) Check for “groupthink.” Another potential problem can arise when no one expresses dissent or questions the moderator or other participants. This was especially relevant to my focus groups because each one consisted of colleagues who were familiar with one another (see 5.2.5), so they might not want to cause any conflicts or they might think similarly due to working in the same organization. To deal with this, I first stated at the outset of the discussion that I was not looking for any consensus but in fact wanted each person’s views. Second, as I listened to the interactants, I noted when disagreements occurred; these were signs to me that the participants felt comfortable enough to express their own views.

Overall, I was very pleased with the quality of the interaction and data produced, as the participants discussed at length the questions I asked and provided relevant answers to my research questions. I then proceeded to set up and conduct the focus groups that would provide the main data for Stage 2.

5.2.5 Arrangement and conduct of focus groups

Through colleagues, I contacted several EAP programs around England; teachers at three expressed their willingness to participate in focus groups. This was thus an opportunity sample, being limited to programs where I had personal contacts. For

each of the three programs, I arranged to meet the participating teachers at an appointed time at their universities; the first two focus groups consisted of current co-workers who were very familiar with each other, while the larger third focus group consisted of postgraduate students who had taught on their school's preessional EAP program at possibly differing times, so some participants were not familiar with one or two other co-participants but were familiar with the rest. As Table 5.1 shows, there were from 3-5 participants in each group; each group lasted between 60 to 90 minutes.

A few points about design decisions need to be made here. First, while the norm for focus group research is groups composed of strangers, it is not a must (Morgan, 1997), and I decided that the advantages of fairly familiar groups outweighed the disadvantages. Practically, I thought it would be extremely difficult to persuade teachers whom I did not know to take the time and effort to travel to a "neutral" location to meet with teachers from other programs for a focus group.

Methodologically, I was concerned that participant familiarity could negatively affect the interaction (see above point about "groupthink"), especially since in each group one of the participants was or had been a teacher-manager, but this concern was allayed by the recognition that familiarity could encourage better interaction, and that the very fact that the teachers were willing to participate showed a willingness and openness to discuss issues; if teachers felt unable to discuss issues with their colleagues and managers, I do not think they would have been willing to participate in the first place.

Second, the small number of participants in the first two groups was somewhat unavoidable and generally advantageous. The small numbers allowed each participant more time to talk and gave me the opportunity to ask more follow-up questions.

Third, three groups were sufficient for my data needs. Krueger (1994) and Morgan (1997) both mention the phenomenon of "saturation"—in which additional data collection does not provide further insight about a topic—often occurring by the third group or so. I found this to be the case: the responses across the three groups

displayed a high degree of convergence and overlap (see 6.2), suggesting that further focus groups would not make a substantial difference to the analysis.

5.3 Conduct of Analysis

5.3.1 Summary and description of Stage 2 analysis

Table 5.2 summarizes the data sources for Stage 2 and the tactics and practices used to analyze them.

Table 5.2 Overview of Stage 2 analysis

Data sources for analysis	
Selective transcripts of 3 focus groups	
Generative analysis tactics	Confirmatory analysis tactics
<ul style="list-style-type: none">noting patterns and themesseeing plausibilitymaking contrasts and comparisonsclustering data & subsuming particulars into generalusing “hidden case”	<ul style="list-style-type: none">triangulating
Main analysis practices	
<ul style="list-style-type: none">keeping notes after each groupcoding of transcript data using MAXqdawriting for external audiences	

Table 5.2 shows that the data I analyzed consisted of transcripts of the three focus groups; these were selective transcriptions because 1) I did not have enough time to transcribe everything, 2) I was fairly certain I could distinguish the data I needed to answer my research questions from the data that was interesting but extraneous, and 3) on one tape, two teachers were nearly inaudible (I transcribed this tape immediately after the focus group so as to recall as much as I could, with the help of the notes that I took during the focus group.).

Compared to Stage 1, I used a more limited set of analytic tactics and practices. In terms of generating tactics, I used ones similar to Stage 1, but with a considerably smaller data set. To confirm findings, I mainly relied upon triangulation. First, I triangulated via data source and looked at whether a finding was supported by multiple informants within and across focus groups and/or by Stage 1 data. Second, I

triangulated by researcher; I gave a colleague with research and EAP teaching experience a list of about 15 codes and asked her to code a section of one focus group transcript (about 100 lines) using the code list and any other codes she might generate. She coded 25 segments; my coded segments covered all 25, and we gave 20 of them the same or very similar code (see Appendix 10 for more detail).

Finally, in terms of analysis practices, I wrote down analytic notes after each group and coded with MAXqda. With the latter, I first coded the focus group data separately to arrive at grounded categories. Then I combined these data and codes together with the Stage 1 data and codes, to aid in data source triangulation. I also conducted analysis as I wrote chapter 6, which presents the Stage 2 findings.

The finding that teachers had some difficulty assessing students' receptive skills (see 6.3.2) can serve as an illustration of my analysis tactics and practices. During the first focus group, a teacher had mentioned that she felt receptive skills were difficult to assess because there was no manifest evidence. I thought this could be a source of weakness in teachers' assessments of students, especially since I was aware from previous reading of the difficulty involved with testing listening (Buck, 2001) (seeing plausibility, noting patterns and themes). Two informants in the second focus group seemed to echo this feeling (triangulating by data source). In MAXqda, I created a code "weakness in assessing receptive skills" under the broader code of "quality of teacher assessments" and assigned this code to the segments in the transcript where those informants had expressed the difficulty. Later, my aforementioned colleague found segments to which this code could be applied, thus lending credence to my own coding (triangulating by researcher).

5.3.2 Issues related to analysis

Focus group data presents special issues in terms of analysis. One issue, which is relevant to interviewing generally, is whether to examine how rather than what things are said; that is, whether to treat group interaction as topic or as resource (Punch, 1998). One could, as Myers & Macnaghten (1999) demonstrate, look at aspects like adjacent turns, back channel utterances (e.g., "yeah" or "right"), and topic shifts, to understand what is occurring in the focus group discussion. Overall, I

think this discourse analytic perspective could potentially provide valuable insights during analysis, but I also think taking on board the assumptions and analytic frameworks⁴ required by such an approach would take the study in a direction tangential to its original and main focus. However, it does remind me to keep in mind some specific focus group analysis principles from Krueger (1994), particularly about considering the context of comments (for example, when the moderator asks an open-ended question, one participant recounts a specific experience, and another participant follows by responding to a narrower aspect of the original question, or when one participant gives extreme comments and another participant gives a response to balance those comments).

Another issue, applicable to interviewing but with an extra twist with focus groups, is the unit of analysis—whether to examine the group as a whole or individuals. I agree with Morgan that:

The attempt to understand a group's activities as no more than the sum of the behaviors of its individual members amounts to the well-known fallacy of "psychological reductionism." The need to avoid psychological reductionism in analyzing focus groups is not, however, a warrant to engage in a form of "sociological reductionism," whereby the behaviors of individuals are treated as mere manifestations of an overarching group process. Instead, we must recognize not only that what individuals do in a group depends on the group context but also that what happens in any group depends on the individuals who make it up. In other words, neither the individual nor the group constitutes a separable "unit of analysis"; instead, our analytic efforts must seek a balance that acknowledges the interplay between these two "levels of analysis." (1997, p. 60)

I think related to this issue is the "co-construction" of interview "texts." (Fontana & Frey, 2000; Kvale, 1996; Silverman, 2001) Increasingly,

interviews are seen as negotiated accomplishments of both interviewers and respondents that are shaped by the contexts and situations in which they take place. ... We are beginning to realize that we cannot lift the results of interviews out of the contexts in which they were gathered and claim them as objective data with no strings attached. (Fontana & Frey, 2000, p. 663)

I agree with this to a degree, but I think interview participants still "bring their own bricks"—that is, they have their own experiences, knowledge, and other resources from which to draw when "constructing" the interview, and these resources can still be investigated.

⁴ Myers & Macnaghten, for example, draw on assumptions and analytic tools from conversation analysis.

5.4 An ethical issue arising in Stage 2

As mentioned in 3.6, most of the ethical dilemmas in this study arose during Stage 1: this is understandable, as video recording—used in Stage 1 data collection—raises a host of ethical issues like what will be done with the videotape. While Stage 2 did not involve such problems, there was one ethical dilemma that in a sense stemmed from, or was “enabled” by, the method of focus group interviewing. In one group, two informants seemed to interrupt each other’s turns somewhat frequently; while this did not seem to prevent either informant from saying what they wanted to say, I did not know what to do with the transcript of the focus group. The informants’ interruptions were ambiguous to me; were they colleagues on good terms but were just used to interacting like that, or was it a sign that they did not have a healthy working relationship, or was it purely a one-off instance? I did not want to exacerbate any existing problem, nor did I want to create one, so I decided not to give that group (or any group) its transcript.

5.5 Summary

This chapter has presented the methodology of Stage 2. It has described the design and conduct of the data collection and the procedure for the data analysis. The findings that resulted from the Stage 2 analysis are described in the next chapter.

Chapter 6 Stage 2 Findings

6.1 Introduction

This chapter describes and discusses the findings from the focus group data collected in Stage 2. I first make some general observations about the conduct of the focus groups, then turn to look specifically at the focus group responses to the research questions. I conclude with a consideration of possible connections between the various findings of Stage 2.

6.2 Some general remarks

Before examining the focus group participants' responses in detail, I wish to make some general observations and reflect on their significance for this study. First of all, perhaps one of the most striking phenomena I noticed was the marked similarity in responses across the focus groups. This is not to say that everyone in one focus group said the same thing; in fact, within a group, each participant often contributed a different aspect of the issue under discussion. However, across focus groups, these different aspects were similarly raised. For example, respondents in all three focus groups believed that teachers gained a great deal of knowledge of students' language abilities through a variety of means; not only that, but the one-to-one tutorial was mentioned in all groups as one of the primary means. In addition, participants in each focus group pointed out the difficulty involved in assessing a student who did not speak up in class. Each group did add facets that other groups might not have mentioned, but generally speaking there was much overlap on many key points.

While this similarity of response can be interpreted in various ways, I interpret it as lending strong support for my methodology. Broadly speaking, it is in line with my critical realist position described in 3.2; specifying my general assertions there, it would seem that these four language programs are operating within social structures that are currently configured in such a way that EAP provision is similar across centers. The possible structures/systems/practices involved are several. Cumming (2003)¹, also noticing a considerable similarity in ESL/EFL writing instructors' teaching and assessment practices, posits that professional networks, conferences,

¹ While Cumming (2001) and Cumming (2003) were based on the same study, the 2001 article reviewed in 2.4 discussed only the assessment dimension of writing classes, while the 2003 article (not reviewed) discussed both teaching and assessment, although emphasizing the former.

publications, post-graduate education, and common discourse and concepts arising from research and theory on second language writing have all contributed to the commonalities he observed. To this list, I would add the university system in the UK (England specifically). It is also possible that, following Becher & Trowler's (2001) idea that cognitive domain influences social organization and structure, EAP provision is at least partly configured the way it is because EAP as a cognitive domain (or group of domains) itself is configured in a certain way. Whatever the influences may be, the result is that teachers in these ostensibly different programs face similar issues and concerns.

Support for my paradigm position also by implication supports more specific aspects of my methodology. One is the appropriateness of my case study strategy in Stage 1: understanding a small number of cases of EAP in-depth can shed light (via analytic generalization) upon a wider range of EAP (and other) contexts because they all exist within or as part of the previously mentioned social structures. Another is the concept of “data saturation” (see 5.2.5) when using focus groups: given that individuals are operating within a common reality (constituted by a range of social structures), it is reasonable to expect that there will be a relatively limited locus of possible responses to a topic—evidenced by the similarity in focus group responses.

Another general observation is that, as in the case studies, all the teachers were familiar with the idea of assessing students in the context of daily pedagogical activity (see 6.3), usually referring to such assessment as “informal,” in contrast to the “formal” nature of summative mid- or end-of-course assessments. This was significant in settling my ongoing concern about the gap between researcher and practitioner conceptions (see 4.3), although at the beginning of all the focus groups, a short time was still needed during which I and the participants had to communicate and become oriented with each other; it seemed to me that the “default” concern of the teachers was often with formal assessment, and it was evident during the discussions that teachers could have said much more about formal assessment if I had given them the opportunity.

One final general observation is that the teachers seemed very positive about being able to discuss assessment in a group. Participants in one group said that eliciting

data through focus group discussion was much better than via a survey, one reason being that the issues discussed did not “fit” easily into pre-determined categories of response. Participants in all three focus groups said they rarely had the opportunity to talk about these things during the course of their daily work: one commented: “[The focus group discussion] is useful for us because it’s given us an opportunity to listen to each other for once; instead of speaking at each other, we’re actually listening to each other.” [G2:145/SL]² While it is unclear as to the reasons why there are so few opportunities—lack of time and the constant press of more urgent matters are presumably at the top of the list—these kinds of comments do seem to indicate that, with willing participants and minimally competent moderating, focus groups involving members of the same organization have the potential to aid in organizational development by creating a “space” for members to talk about professional issues. More importantly, at least for this study, comments like the ones above add further confirmation of my methodological choice of focus groups as suitable for collecting relevant data that will allow me to theorize about practice, one of the general goals of this dissertation (see 1.3).

In the following sections, I examine the focus group responses in an attempt to answer the research questions posed.

6.3 RQ5. How do English language teachers think they can increase the quality of their impressionistic knowledge of a student’s language abilities?

6.3.1 Teachers’ confidence in their judgments

Before looking at how the participants thought their impressions could be improved, it should be mentioned that they generally expressed a high level of confidence in their knowledge of students’ language abilities. This confidence was founded upon the nature and amount of information they gained about students over a class term [see for example G1:34/TT; G2:6-10/SL,12-14CC; G3:5-7/RB,33-35/RC,51-53/LB]; teachers drew upon their observations of student-student interaction in the classroom, student-teacher interaction in and out of the classroom, and products like

² Focus group data are referred to by [GroupNumber: TranscriptLineNumber: Speaker] (this last is not used when speakers are indicated in the excerpt itself).

presentations and essays (particularly when drafts were involved), and these offered a “huge” (this adjective was used by participants in two groups) amount of data. This was also sometimes supplemented by “formal” data from placement, diagnostic, and/or mid-course achievement tests.

It came out in the focus groups that one particularly rich source of information was the regular tutorial sessions—usually at least 15 minutes per weekly or biweekly session—between teachers and students [G1:34/TT; G3:29/RA,146/LB]. This did not come up in the case studies because the observed insessional courses did not offer tutorials; however, the presessional courses of the Stage 1 and Stage 2 language centers usually did. One reason tutorials were such a rich source of data was not only because the topic of discussion was a student performance (usually a writing draft), but the act of discussion was also a kind of spoken performance [G2:32/SM].

Another reason was the tutorials’ one-to-one nature:

For me, the tutorial’s certainly where I do the main teaching, the thing where I get a student one to one and see what they’ve done, and say this is good, it could be better if you did this and you still need to work on this and keep doing this, and seeing how they’re reacting to it. That’s where it really happens and that’s where I get to build students up if they need building up. The very public form of the classroom is difficult for encouraging people or dealing with any difficulties. [G3:79/RA]

This last comment also lends support to the idea proposed in 4.4, that the size of assessment “audience” influences teacher thinking (see Figure 4.1).

I think it is important to mention that while focus group discussion about judgments of students centered on the impressionistic knowledge used for formative purposes such as giving feedback to students and guiding classroom management, the teachers also extended the aforementioned confidence to the summary judgments they made about students at the end of a course (this could involve qualitative comments and/or a level grade). Most of the participants said that such teacher assessments were as good as or better than the IELTS in determining a student’s ability to succeed academically—a view also shared by CTA in Stage 1 [A1:41-46]. The main arguments were that 1) IELTS was like a snapshot taken at one time that could only provide broad levels, while teachers’ assessments occurred over a period of time and could give finer descriptions of a student’s abilities [G1:15/TM,45/TT; G3:154/RA]; this was also CTA’s argument (see 4.2 and note 1 in chapter 4)]; and 2) students

could be trained specifically for the IELTS (the implication being that this did not reflect students' actual academic abilities) [G2:142/SL; G3:159/RC].

In response to prompt 2 (see Appendix 9), in which a Stage 1 teacher recalls how he made an incorrect judgment of student ability before grouping students, and the question of whether they misjudged students' abilities, the teachers did admit that mistakes were possible and even common [G1:22-23/all; G2:128/CC,129/SL]. As illustrations, one teacher described how he had a student who did not regularly attend class and did not interact much with her classmates; this had led him to believe she was relatively low level, until she came to his office carrying a National Geographic book and asked about the meaning of the words she did not understand, which were actually quite advanced. [G1:25/TT]. Another teacher recounted:

CC: I had an experience last summer of one Taiwanese student ... who the minute she walked into that class she was talking, she was talking fluently, she had very good pronunciation, American pronunciation, and she's contributing freely, and I immediately thought oh right this lady is obviously very confident and very fluent and very good at listening, and subsequently I realized that she was all of those things but she wasn't actually that good at listening; she was just giving a kind of... she was starting off with her strongest...

SL: that was her strategy

CC: that was her strategy; she wasn't weak but she wasn't as strong as I'd judged in that first five minutes. [G2:133-135]

However, while the participants did feel such misjudgments of ability were common, they generally (see an exception below) also felt this often did not matter so much, at least for the purposes of classroom management like grouping, compared to more pressing considerations like class dynamics, student personalities, and creating a lively lesson [G1:26-28/TJ&TM; G2:129-130/SL&SM]. Besides, as the above excerpt suggests, teachers felt any misjudgments will likely be corrected over time as more information arises; this echoes an argument made by Harlen & James (1997) about how formative assessment becomes more reliable over time.

6.3.2 *Areas of weakness in teachers' impressions of students*

While the teachers expressed confidence overall in their impressionistic judgments of students, the focus group discussions did highlight two significant areas of weakness. One was the assessment of receptive skills generally and listening in particular. A participant in focus group 1 said teachers often think they know a student, but it was difficult with listening and reading ability. “even with the most careful of assessments,” essentially because these abilities occurred in the student’s head and were invisible to the teacher [G1:21/TM]. A group 2 participant echoed similar sentiments: “In classroom activities, if they’re doing listening activities, they don’t necessarily have to say or write anything but they may actually have understood everything that was said to them. So I mean you don’t actually have any evidence to judge them on their listening or indeed on their reading.” [G2:49/CC]

One teacher suggested that an aspect of this problem of assessing listening was, in effect, construct under-representation (or lack of content representativeness) (Popham, 2002):

What we have done in the past is that all the teachers who teach students—and it doesn’t matter whether they’re teaching a particular student reading or writing or listening and speaking—have given assessments of listening skills [In this program, teachers gave assessments of students that were at least in part based upon their impressionistic/informal knowledge of students; this illustrates the point that information teachers gain about students can be used for a variety of purposes—for feedback, classroom management, end-of-course reporting, or otherwise.] on the basis of how students respond to the teacher speaking or the other students speaking in class. I personally have a bit of a problem with that because I don’t think it’s the same kind of listening that we’re really preparing them for; the main kind of listening that we’re preparing them for is listening to lectures and taking notes, and for me that’s a different kind of listening. ... What I’m saying is if you assess listening on the basis of how students appear to be understanding what’s going on in the classroom situation—and it could just be some group work or something like that—then I think that’s not the same as actually listening to a lecture. [G2:23,25/SM]

Another important source of weakness was misjudging the activeness of the student. This took two forms. First, a student who was quiet or silent in class was difficult to assess; the participants pointed out that such silence can be interpreted in different ways—something I term “the ambiguity of silence”:

CC: I mean we’ve had this ongoing discussion and doubtless will until we’re all dead about if you do have a seminar, part of participating in

seminars is actually saying something, but if people don't say anything, what does that mean? Does that mean that they're not actually able to follow what is going on? They may very well be able to follow what is going on. Does that mean then that they can't express an opinion? No, it doesn't. It may mean that for some reason they choose not to. Maybe they don't have an opinion.

MY: You can't see it.

CC: Exactly. So you have some evidence, but if the evidence is silence, then that could mean so many different things. [G2:53-55]

TM: [responding to prompt 2] ...If you are assessing students on observations of classroom behavior, this example sort of highlights that it is dangerous. I am just thinking as the mother of a girl, not an English language learner but who has made a positive decision never ever to speak in class, to go through a whole school year without opening her mouth but when she has to... it's very hard to assess someone like that. ... It's very complicated. [G1:33]

This ambiguity could perhaps lead to an underestimate of the student's abilities [G3:72/RA].

Conversely, a teacher might overestimate the abilities of a student who was talkative and active in class [G3:76/RB; see also G2:133-135 illustration above and G1:77/TM excerpt in 6.5.1 below].

6.3.3 *Ways of improving teachers' impressions*

Of the handful of ways the focus groups suggested to improve teachers' impressionistic knowledge of students, two were related to the problem of misjudging students' activeness in class or lack thereof. One participant suggested using texts from a variety of subjects when planning classroom activities; this would give students who were perhaps quiet because they were unfamiliar with a particular subject of a text an opportunity to talk about a more familiar topic [G3:65-69/RC]. Along a similar line, one participant suggested always changing student groupings so as to make sure students were not overly quiet or talkative due to their interaction partners: "This student might be able to work with that student and they don't get on, or he talks too much and this one doesn't get a chance. ... So I try to put different people with different partners; then you can see a different side." [G3:74 LB]

One possible solution to the difficulty of assessing receptive skills was using formal assessments of those skills [G2:28/SM]. However, there were limits to testing. As

one participant said, teachers wanted to monitor progress but tests often worked against the idea of development [G1:39/TM]; besides, tests could not—or should not—take up too much time:

SM: As I said before, my main assessment of listening is on two formal tests during the term but two really... I've just said that there's lots of kinds of variables; to have a reliable assessment I would really have to base it on 10 or 20 assessments but you've got to spend some of your time teaching and some of your time ...

CC: You've got to start the lesson. [G2:132-133]

Another solution was that, while making judgments about students at the very beginning of a course was perhaps inevitable, teachers needed to be wary of those initial impressions [G2:128,137/CC]. This simple warning may take on somewhat greater importance when one considers that two of the four examples given by teachers of misjudgments (three from the focus groups and one from Stage 1, which was used for the prompt) involved initial diagnosing of student ability.

6.3.4 Discussion: dealing with blind spots

The participants' general confidence in their knowledge of students based on a wide range of sources and a large quantity of information echoes the arguments, mentioned in 4.8.1, that have been put forth by proponents of teacher assessment (Huerta-Macias, 1995; Harlen & James, 1997). Against this, as also mentioned in 4.8.1, less enthusiastic writers point out that the quality of such knowledge still cannot be assumed (Brown & Hudson, 1998; Popham, 2002). Considering these arguments and also the Stages 1 and 2 data, I would take something of a middle position: while teachers can have a great deal of information about students, there may be blind spots—for example, the Stage 1 data suggests that a teacher's pedagogical approach influences the nature of the information teachers gain about students, and the Stage 2 data suggests that receptive skills are not easily assessed—that need to be recognized. On the other hand, while formal tests of particular skills can provide more systematic data, they too may have blind spots, possibly in the form of construct underrepresentation or irrelevance, etc.: in reply to Popham's comments that the conclusions he had drawn about his students based on his observation-based judgments as a teacher were far too generous (see full quote in 4.8.1), I would say it could be that the conclusions drawn from the exam were far too

strict. The way forward, it seems, is to consider how testing and test results can be usefully integrated into the mental portraits that teachers have of students in order to improve decisions made (such as for feedback, lesson planning, or end-of-course reporting) without adverse effects upon non-testing time (i.e., time taken with testing must be time well-spent).

In sum, the focus group data showed that teachers had a great deal of confidence in their judgments of students, but that there were clearly aspects that were effectively “blind spots,” like students’ receptive skills, and that could be improved. The responses to the RQ5-related questions emphasized the individual teacher and his/her experiences in the classroom. Responses to RQ6, however, focused on the organization and its programs; these responses are described and analyzed in the next section.

6.4 RQ6. How do teachers who have experienced managerial changes like a more explicit syllabus and more explicit summative assessments feel such changes have affected their assessment practices?

Somewhat to my dismay, I came to realize that the “assessment practices” part of the question could not be adequately answered by the focus group method as I carried it out. One problem was that because I had conceptualized assessment (see 2.2.2) to include observation in pedagogical activities, I found myself asking about how such managerial changes had affected their teaching more generally rather than their assessment practices in particular (since they were intertwined in my conception). This in turn led to responses that discussed general changes in teaching rather than specific changes in assessment practices. In fact, I reached the conclusion that from a methodological perspective, the best way to answer the question as originally posed would indeed (as I had previously thought; see chapter 5 note 2) have to be to find teachers who were about to go through such managerial changes and collect both baseline and post-change observation data (much like how Wall & Alderson, 1996, studied washback in Sri Lanka).

My dismay at not being able to answer the assessment practice issue—and thus being unable to directly confirm my “plausible hypothesis” (see 3.5.2) from Stage 1

that macro-level managerial changes would affect micro-level informal assessment—was offset by the fact that the wider question of how managerial changes affected teaching generated fascinating responses.

First of all, two of the three programs involved in the focus groups, like the Stage 1 language center, seem to have gone through a similar historical process.³ At some point in time, often during the 1980s and early 1990s, teachers at these centers had had a great deal of freedom to choose their courses' content and assessment methods. However, over a period of a few to several years, that freedom became more restricted; course content became more prescribed through a more explicit syllabus and summative assessments became more uniform [G1:39/TM; G2:112/SL].⁴ In Stage 1, CTA had already given a few reasons for the changes at his center: to encourage other departments' acceptance of the center's assessments, to align course assessments more closely with the kinds of assessments students would face in their departments, and to withstand external examiner scrutiny (see 4.2); in Stage 2, one participant spoke about why these changes had come about at their center:

... When I first worked here [in the late 1980s] things were much more ad hoc in terms of course materials and testing and assessment, and in fact very much dependent on the whims I think of particular course directors or even of individual teachers. Gradually as time has gone by it became very clear that this was detrimental in many ways and there was a need to be prescriptive both in terms of course content and syllabus, timing of delivery, and ultimately of assessment. And there are a number of reasons for this. One of them is the very fact that a lot of our summer teachers come from all sorts of backgrounds, and although we check them up very rigorously as much as we can, we're relying on references and interviews as much as possible, we never really know what we're getting until they come. And therefore we kind of impose this prescriptive approach simply because of the pressures that *** [another participant] mentioned, because let's imagine we get a student in at the beginning of July IELTS level 5.5 and we know that in October that student is going to be thrown in with a bunch of British kids, and they're going to be exposed to a really huge amount of reading—literally hundreds of pages of reading in a week in a foreign language—and they're going to be studying a complex subject, and therefore we really have to use our experience and our knowledge to channel the efforts of the students to get up to that level as much as possible without too much digression. ... [G2:112/SL]

³ The third focus group consisted of postgraduate students who had taught on their university's preessional EAP program; none of them had had more than two or three years' experience with that program, so were not in a position to comment on the program's historical development.

⁴ This phenomenon seems to have gone unnoticed in Jordan's (2002) overview of the development of EAP in Britain, although it could be seen at least in part as a response to trends he mentions, like the expansion of EAP, which would perhaps require a more systematic teaching and assessment "technology" to handle the increasing number of students. This could also be seen as a program-level instantiation of broader managerial trends in higher education (Deem, 2004).

While there were thus desirable goals behind making the syllabus more prescribed and the summative assessments more uniform—to encourage consistency among teachers, better prepare students, etc. (participants in focus group 3 also saw the value of these goals [G3:123/RA,125/LB]—several participants expressed feeling a tension between these institution-/organization-wide goals and constraints on one hand and the desire to help students as individuals with their individual problems on the other [G2:112,114/SL,109/SM; G3:145/RB]. This tension seemed to manifest itself most strongly as a pace problem. That is, the organizational requirement to cover prescribed syllabus content in a fixed amount of time—with some of that time perhaps being taken up with mid-course formal assessments—implies a certain set pace at which teachers must work to “get through the material.” However, this may pressure and rush teachers who feel their students need more time on a topic [G1:48/TO]; in addition, because students are at different levels, a uniform pace may hinder some students’ learning [G2:121/CC].

I think it is significant that the participants’ responses in this area focused particularly on the impact of a more explicit syllabus; this was also the source of pressure for Teacher CTB in Stage 1, whose comments were an important reason for my asking RQ6 in the first place. This is not to say that an explicit syllabus in and of itself was a problem. One teacher said that when she was new to teaching a particular course, she found its fairly detailed syllabus extremely helpful, giving her a good idea of what to focus on [G1:38/TJ]; another teacher felt that while there was pressure from the syllabus, it also provided structure and kept the class moving [G3:145/RB]. It seems, rather, that an explicit syllabus is only a problem when a teacher perceives it to be overly detailed or prescriptive in terms of content or timing, particularly in relation to the level of the students in the class; if so, then time pressure and pace become an increasingly weighty element of teacher thinking in class and during planning. (As I mentioned earlier, however, whether this concern actually affects teachers’ classroom assessment practices remains a viable but as yet unanswered empirical question.)

Intriguingly, one participant proposed that teachers’ perceptions of the syllabus partly depended on their view of their relationship to their organization:

There are some teachers who view themselves as part of the team and they're all working together to meet uniform objectives. There are other teachers I've met who believe when they're in the classroom, they do whatever they want to do and nobody else is going to tell them what they should be doing. And their view of the institution is a little bit like a sort of hairdressers' franchise, where you rent a premises and you might have 4 or 5 different hairdressers working in the hairdresser's and they've all got their own business but they're sharing the costs of the premises and that kind of thing. [G2:115/SM]

If this is true, then in light of the aforementioned trend towards uniform syllabus and assessments, there may be less and less diversity in teaching (and thus assessment) methodology as there is less and less room for teachers to “do whatever they want to do.”

Focus group participants did not express any special concern about uniform summative assessments (although, as noted above, mid-course formal assessments for formative or summative purposes were a concern if they cut into teaching time). The exception was courses offered by the language center that prepared students to take exams like the IELTS or CAE (Cambridge Certificate in Advanced English); a teacher on such a course would likely feel a lot of pressure, so much so that, as one teacher said, “it would literally be [about] getting them through the hoops.” [G1:36/TJ]

This finding suggests that there is no strongly perceived “micro-washback” effect posed by uniform summative assessments, as I had originally conjectured (see 4.8.2). There may be several explanations for this. One is that teachers may believe such uniform assessments are necessary so that teachers will calibrate their judgments with each other:

... We all know that these students we are teaching in different groups are actually going to go on to maybe even the same academic course from 5 or 6 or 8 or 20 different groups, so the people receiving those students need to have a report which is not just my subjective assessment or his subjective assessment or someone else's. There has to be standardization. ... [G2:119/CC; also G3:123/RA]

Another explanation is that—unlike government-mandated exams, which the term “washback” usually refers to—the summative assessments on EAP courses, at least at these centers, have a high degree of teacher involvement in their construction, administration, and scoring. It may be—and this is purely another emergent

“plausible hypothesis” to be researched—that as teachers have more power to make changes if dissatisfied with some aspect of the summative assessments, they feel less imposed or intruded upon in their classroom teaching by those assessments.

6.5 RQ7. Do teachers use stereotypes and projection when they conduct interactive assessment of students?

Based on the data, the basic answer to this question is “yes.” In fact, outside of one participant’s answer about projection, the responses to this question were fairly unanimous. However, the responses also delineated these concepts more clearly, so as to add some complexity to the two concepts as I originally proposed them. Stereotypes and projection are dealt with separately below.

6.5.1 Stereotypes

When given the prompt and asked the question about whether they had preconceived ideas about aspects of students’ language ability based on their ethnicity or nationality, the participants gave remarkably similar answers.

Overall, participants affirmed that they did use stereotypes; one teacher expressed this particularly clearly:

Well I think we all have preconceived ideas—maybe I shouldn’t speak for everybody!—but I do. I know what to expect from *** speakers, and I know what to expect from Japanese and Chinese and Thais ... in terms of the way they’re going to express themselves, the level of accuracy, the depth of their content I’ve been teaching for a long time, and I’ve always for all of my career dealt with students from other countries, so I have without a doubt formed opinions, preconceived ideas of virtually every national group, certainly continent-wise anyway, even to the extent of people from the States and Canada and what have you. I can almost predict in a way ... what they’re going to say by how they’re going to say it. [G1:86-90/SL]

Participants also mentioned examples:

What this teacher [in the prompt] described about the Italian student who seemed very good speaking fluently but noticing he had problems with his analyzing text, I can think of many cases when I’ve seen that—or you get an Italian student’s writing [and it is error-filled]—the student’s fluent, speaking so fast that you didn’t notice the errors. [G1:77 TM]

But I have to say I have the same stereotypes myself. I certainly agree with [SL]. If you're talking about spoken language, Japanese students appear to lack fluency, they appear to mentally rehearse what they say before they say it, whereas Arabic students place more emphasis on fluency and less value on accuracy. [G2:99/SM]

But the teachers were quick to point out that the terms “preconceived ideas” or “stereotypes”⁵ were inaccurate in that they denoted a lack of knowledge, when in fact teachers had a great deal of experiential knowledge about different groups of students. As one participant said, “stereotyping is a very negative word. The positive form of that is actually having a wealth of experience to know how to deal with [students of particular backgrounds].” [G1:80/TJ] She added that stereotypes usually involved faceless people, but it was different for teachers when faced with a student and a name. Likewise, another participant said, “‘preconception’ sounds like such a negative term ...; [it suggests] making [judgments] based really not on a lot of data, whereas something like what [SM’s] saying is that actually we have had a lot of experience. So they’re obviously generalizations but they’re generalizations which I think can help us” [G2:93-97/CC] These generalizations could also be derived from books [G2:98/SL; also see sources of assessment cognitions in 4.6].

The main benefit of using what I have termed stereotypes was that teachers could focus students’ attention on particular problems commonly faced by particular groups of students. For instance, one teacher said:

one thing that I say quite a lot to the Chinese speakers is Chinese doesn’t have syllable-final consonants. Therefore the Chinese speakers do this or this. ... If a student knows what’s happening in the first language and what happens in English in a very general sort of way, [it gives them something to work on].” [G1:81/TO]

Another benefit was that stereotyped information could help a teacher develop rapport with students through showing them that the teacher understood their cultural and educational background while at the same time pointing out aspects of British academic culture that needed to be learned [G2:93/CC].

⁵ I was careful to word the question using the former term, but participants used not only the former but the latter, without my prompting.

As the previous data extracts [also G3:109/LA] show, the stereotypes teachers had were usually of a linguistic nature (e.g., pronunciation, fluency, accuracy, lexis). However, this also could extend to assumptions about students' knowledge of academic expectations, as with attitudes towards plagiarism [G3:112/RB], and about students' educational and cultural background, which one teacher felt was dangerous because there was an assumption of student homogeneity within national or ethnic groups that may simply not exist [G2:99/SM]. A colleague responded to this by saying that, besides the fact that there did seem to be a common core culture among students of a group around which they still varied,

anyone who's been through an academic background I think is not going to fall into the trap of stereotyping every single person from a country as having exactly the same prior experience or react to that experience in the same way. It's helpful but of course we're not prisoners to that kind of knowledge. [G2:102/CC]

And this is an important point to keep in mind. I have used the term stereotype to label the cognition described in the Stage 1 findings primarily because it captures a sense of how teachers' knowledge about particular ethnic or national groups of students seemed to be "stored" in the mind and employed in assessment. I still think it is a useful way to look at this kind of knowledge, always keeping in mind it is in a neutral sense. However, at the same time it must be remembered that it seems these stereotypes only serve as a kind of aid to inform teachers' instruction generally and assessment specifically. They were neither static—"I would be lying if I said that before I taught specific ethnicities that I didn't have specific preconceptions, but if you get feedback that says otherwise, you incorporate that and you change your expectations about different groups and so on." [G3:110/RC]—nor did they prevent a teacher from "seeing" the student as an individual with individual characteristics, linguistic and otherwise:

obviously, you're going to be careful. Hopefully on one hand you look at students completely openly and honestly, but having experiences with what you know of certain cultures is going to affect positively the way you deal with a student in contact hours. Sometimes there's surprises and that's great... ." [G1:76/TJ]

6.5.2 Projections

In Stage 1, it was found that when assessing a student performance, a teacher may imagine or “project” that student’s performance into what can effectively be described as a target language use (TLU) situation, such as a classroom discussion or presentation involving British students (see IV under 4.5.2). The Stage 2 data to a degree confirmed that this happened, but the visual aspect I originally proposed was not mentioned.

Most of the participants agreed that they often thought about the TLU situation when assessing students and giving feedback:

We’re constantly predicting, projecting students, into academic situations. ... A student’s whose always [quiet in one’s EAP class], if they’re on an MBA [Master of Business Administration degree program], how are they going to cope with a whole load of very articulate, very pushy [classmates]? [G1:57/TM]

It’s always in mind...you’re preparing them for the bigger wider world. ... In tutorials ... you’re constantly saying to them, you’re constantly talking about their work in the context of what they **will** be doing.” [G1:60,65/TJ]

I particularly think about that when I’m writing reports and making decisions, say, as a course director about the final grade that’s going to be passed on to the departments. You’ve got to think about that realistically both for the benefit of the individual students and also for the people they’re going to join. [G2:57/SL]

It’s got to be always about the end product where they’re actually going to have to use language, what kind of context, you got to be thinking about that. [G2:58/CC]

To generalize a bit, I see [assessment] as an informal and really broad data collection. The only concrete thing I have in mind is what they will be asked to do [for their courses]. [G3:33/RC]

According to the participants, teachers derived their knowledge of the TLU mainly from either personal academic experience or from talking with students who may have taken a preessional and were currently in the midst of their degree program:

In fact, we know the departments, we know what students are like. We work on the insessional program, you know the [learning] environment that the students will be in [G1:69 TM]

Sometimes—fairly frequently actually—I meet students on the campus after they’ve left us ... and chat [about] what they’re doing and how is it and is it what they expected. So you do pick up snippets. Also of course we use our own personal experience; we’re all academics and we’ve spent our whole life in academia.... [G2:69/SL]

I draw on my current continuing experience as a student to put myself in their shoes, and I think what would I do in this situation, how would I approach this task, not necessarily the right way but as a way of doing it. And I think that applies particularly when they do the parts of the course where they go into real lectures and seminars and interacting with other staff members of the university who are not directly involved in teaching the course. I can really identify with that because that’s the same position I’m frequently in. [G3:81/RB]

This last point echoes CTA’s comment in Stage 1 (see IV under 4.5.2) that he drew upon his own experiences of being an undergraduate and later a master’s degree student in classes with overseas students.

In addition to the sources mentioned above, one of the participants was conducting an empirical study of the language demands Asian students actually faced during their time of study at their institution; this seemed to give that participant a much more detailed picture than he would otherwise have [G2:68-69/CC].

While this kind of detail may be available, teachers still seemed to “project” the student into a kind of general academic situation:

CC: ... we don’t always know what the individual context is going to be like on any one course in any one department.

SL: We generalize.

CC: You get lectures next door with like 200 students; no one’s really expected to ask questions at the end of lectures. You get other lectures where you got 15 or 16 students; they are expected to respond immediately to questions thrown out in the middle of lecture, so the actual demands on them can vary quite a lot. ... So the kinds of things that we’re thinking about are generalized into a [kind of] artificial as well as common denominator academic context [G2:58-60,66; also G3:98/RA]

This meant teachers often focused on form more than content:

LB: I think it’s hard to visualize say a science student because I don’t have current experience of being a science student. I can deal with [students in fields similar to mine], but because I’m not a physicist, [if] there is a potential PhD physics person giving a presentation, for me that would be quite hard, to reach a hard judgment on that.

MY: So if you’ve got that kind of student in class, how do you judge?

LB: I think I do it subconsciously; I think I tend to try not to get carried away with the content of what they're doing, because I'm not going to follow that through and I'm not going to understand too much about that. My job is to assess the way they do it, the attitude, even things like the way they dress and the way they stand, because a lot of that can be culturally bound as well; it can give off wrong messages. And I think part of what I do is to try and get people to give off the right message. ... I find myself talking about making eye contact, taking their coat off, standing straight, not overdoing the gestures, all sorts of things; I think those things are quite important. [G3:86-88,89]

As something of a counterexample, one teacher argued that this generality was just “common sense” and did not require knowledge of the TLU:

There are things which I think you should and shouldn't do in a presentation which are kind of common sense, but because students are in a slightly stressful situation they can forget it—having your back towards the audience—it's common sense that your audience should see your face. It's common sense that you should not read out verbatim what you've got written on the slides. Why read out something aloud which your audience can read for themselves? So they're kind of common sense things; I think it's just good presentation technique. It's the same whether it's an academic context or a business context or a training context. And then there are other things which are related to that—grammatical accuracy is important, accurate pronunciation is important, considering the background knowledge of your audience, what terminology they're going to be familiar with, what you're going to have to explain to them, organizational things... I have to say I haven't a clue what goes on in presentations given in academic contexts in [this university]. [G2:76/SM]

Therefore, it may be that this generality is more relevant to oral presentations, while written work would require more specific knowledge of academic expectations.

It also seems that fairly early on in the class, teachers look out for extremely weak students and become concerned about their passing the requirements for the course [G1:56/TT]. This process may involve projection—TT's comment arose in response to the question about projection, and CTB from Stage 1 seemed to be predicting SJ's future performance in the class:

B: SJ ... seemed to have quite a bit of difficulty forming an argument; sometimes it's hard to judge, but it may be the question that he had chosen. But even so, it's not such a worry as to imagine that he's going to drop below the class; he may be in the lowest band or he may be in the second band up, so there's nobody that I'm really worried about in the spoken aspect. [B3:53]

Oddly, there was a lack of any response as to the visual nature of projection. I asked the focus groups about 1) whether they imagined their students in a TLU' situation

and how they would be assessed, and if so, 2) whether this imagining was visual in any way; however, the focus group teachers only responded to the first question, not the second. This lack of data leaves it unclear as to whether there is a visual component—perhaps it really does not exist, or it does exist but evidence for it can only be found through a method like stimulated recall, where thoughts are remembered in concrete contexts.

6.6 Some concluding comments on Stage 2 findings

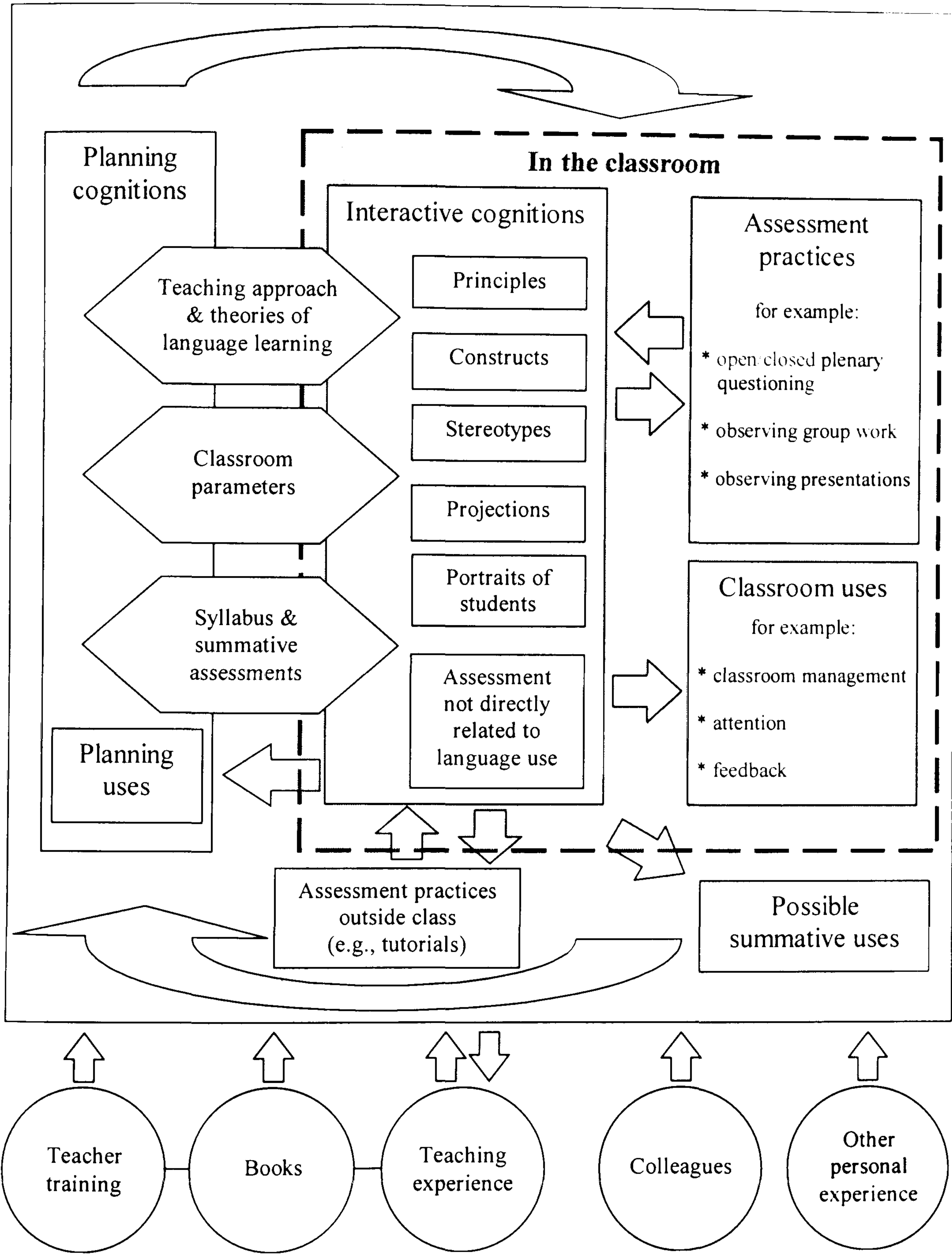
The findings described in this chapter confirmed and elaborated many of the findings from Stage 1. Overall, the Stage 2 findings confirmed the Stage 1 finding that teachers gain information about students and build up a mental picture of a student through an array of assessment practices; the Stage 2 data expanded the assessment practices to include not only observation during classroom activity but also observation outside class, such as during tutorials on student products. The Stage 2 findings also highlighted areas of difficulty for teachers, such as in the assessment of listening ability, and possible means for improvement. In addition, the Stage 2 findings not only confirmed stereotyping and projecting as cognitions in teachers' minds as they assessed student performance, but also elaborated on their sources and nature.

The Stage 2 findings also led to the hedging or complete revision of other findings from Stage 1. For instance, while projecting in some sense was recognized by teachers, there was no data to confirm or disconfirm that this projecting was visual in nature. Also, the focus group data showed that what was proposed as a relatively simple relationship between increasing managerialism via explicit syllabi and uniform summative assessments on one hand and classroom assessment practices on the other was not only much more complex but also very difficult to empirically verify.

At first glance, the answers to the three research questions posed in Stage 2 seem disparate; one deals with quality of informal impressions of students, another deals with managerial action, and the third deals with particular mental activities. However, using Figure 6.1 (the same model of teachers' assessment cognitions

network presented in Figure 4.2 but slightly amended to include assessment outside class (see 6.3 above)) as an aid, some connections between these three can be drawn.

Figure 6.1 Amended model of teachers' assessment cognition network



First, it seems that managerial action can have an impact on the quality of teachers' informal knowledge of students; an overly explicit syllabus, for example, may through time pressure force the teacher to spend less time with assessing students and more time covering material. This is not to say it would change a teacher's

teaching and assessing practices, as I had previously hypothesized, but it may mean that those practices become cut off or abridged because of time, as indeed was mentioned in the stimulated recall data of Stage 1. Also, by influencing assessment practices in this way and in other ways—for example, summative assessments leading teachers to prepare “mock” assessments in class for formative purposes—managerial action can affect the amount and nature of information teachers gain about students.

Second, as the model proposes that stereotypes and projection are in interaction with assessment practices, which in turn contribute to teachers’ knowledge (mental portraits) of students, it may be possible to strengthen the quality of teachers’ assessments by strengthening the knowledge base underlying these two cognitions. While the research literature, needs analysis studies (like the one mentioned by one of the focus groups), and other resources can help, it would seem from the focus group data that because stereotypes and projections are largely built upon personal experience, whatever knowledge is gained through these channels needs to be put through practice in the classroom before teachers can incorporate them into their day-to-day assessment cognitions. Related to managerial action, the organization as a structure can perhaps facilitate this process in the interest of professional development.

6.7 Summary

This chapter has set out the findings from Stage 2 of my study. After some general remarks about the focus groups, I present the findings on improving teachers’ impressionistic judgments of students, the influence of managerial actions, and the nature and role of stereotyping and projection in teachers’ assessment cognitions. The chapter concludes with a discussion of how these findings relate to each other and to the Stage 1 findings.

Having concluded the empirical portion of my study, I look in the next chapter at the findings as a whole and how they advance knowledge of the field, and I consider recommendations for research, policy, and professional development.

Chapter 7 Discussion & Conclusion

7.1 Introduction

This dissertation concludes with 1) an overview of the completed study and its findings and 2) a discussion of this study's contribution to the field and its wider relevance. I first briefly summarize what I did in this study and then review the study's main findings. This is followed by a discussion of the limitations of this study. I then integrate my findings within the larger picture of research done in this area to date, explaining how it confirms and extends current knowledge. Finally, I make recommendations for research, policy, and professional practice.

7.2 Summary of the study and its findings

In this study, I have explored the cognitions underlying EAP teachers' classroom assessment practices and issues related to them. In Stage 1, I began with some general questions about the nature of teachers' assessment practices, the mental activities behind those practices, and the sources of those mental activities. I researched these questions via a case study strategy, looking in-depth at two teachers of an insessional EAP course at a UK university language center. Data was collected primarily through classroom observation and stimulated recall; this data was then analyzed, answering not only the original set of broad research questions but also generating three further and more specific research questions (see chapters 3 and 4). In Stage 2, these questions were answered through analyzing the responses of participants in three focus groups comprised of EAP teachers at two other UK university language programs (see chapters 5 and 6).

Taken together, the main findings from Stages 1 and 2 can be organized according to the three key accomplishments of this study:

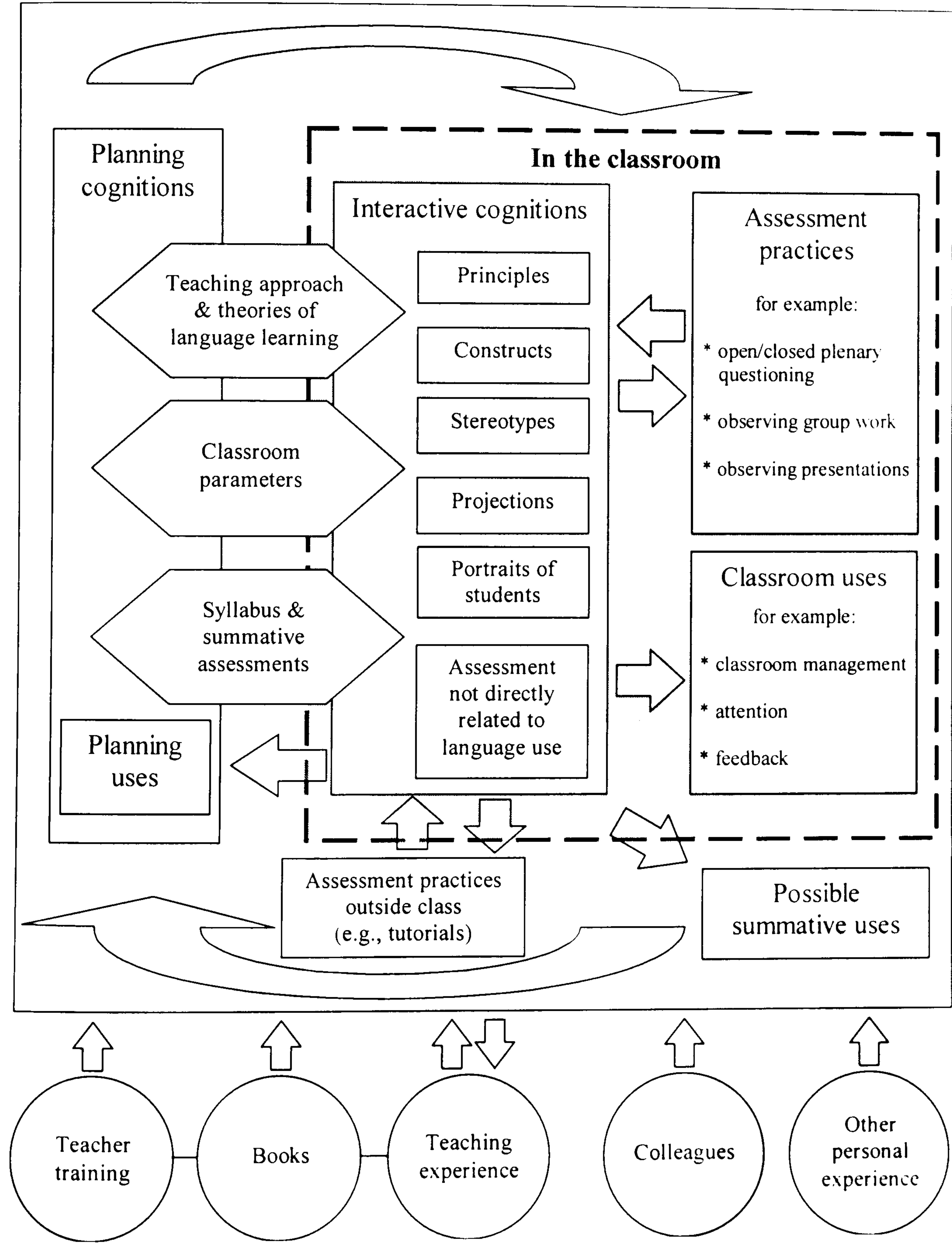
- 1) mapping the components of what, for heuristic purposes, can be called a teacher's assessment cognition network;
- 2) identifying areas of possible weakness in this network;
- 3) exploring the influence that the educational organization exerted upon this network.

These are discussed below.

7.2.1 Components of the assessment cognition network

This study found that when teachers assessed students in the classroom, a wide array of mental processes, resources, and considerations were involved. These interrelated cognitions can be described as comprising what may be called an “assessment cognition network,” depicted visually in Figure 7.1.¹

Figure 7.1 The assessment cognition network and its sources (same as Figure 6.1)



¹ This repeats Figure 6.1, which was an amended version of Figure 4.2.

According to Figure 7.1, the major influences in this network, mainly during planning, were personal teaching approach and views of learning, syllabus and summative assessment requirements, and classroom parameters such as class size and timing (see 4.5.1). These “strategic cognitions” pivotally shaped or “set the table” for the kinds of assessment practices that were used. They also had some influence upon teacher thinking during interactive assessment, i.e., assessment that occurred during class time.

During such interactive assessment, another set of “interactive cognitions” in this network came into play (see 4.5.2). These included:

- assessment principles—Teachers had a personal set or credo of maxims, centered on ideas of fairness and student benefit, which guided their assessment practices and use.
- constructs—These were what teachers looked at when they assessed students in the classroom. Understandably, they focused on linguistic aspects like grammar and pronunciation; however, teachers could also look at non-linguistic aspects like critical thinking ability.
- projections—Teachers imagined how others in the target language use situation (in this case course tutors and native-speaker classmates) would assess the student’s performance, then made judgments and gave feedback accordingly. It was unclear whether this projecting was of a visual nature, but it was apparent that some kind of comparison of student performance with what teachers deemed to be target language use performance standards did occur.
- stereotypes—Based largely on teaching experience, instructors had knowledge about common features (particularly linguistic but also cultural) of particular national or ethnic groups, and this knowledge seemed to shape their judgments and the feedback they gave to students.
- mental portraits of students—As teachers assessed students via assessment practices, they began to build up impressionistic portraits of their students. These portraits, painted in broad strokes in early assessments, not only informed later assessments but were also filled in or perhaps significantly corrected by them: they were thus of a dynamic and iterative nature.

The information about student language ability gained through interactive assessment was used in several ways, mainly for pedagogical goals in this study (e.g., classroom management, feedback, and lesson planning), although teachers sometimes used their impressionistic knowledge of students for summative assessments also (e.g., to mitigate a student's poor performance on a summative assessment; to provide qualitative data on students' reports to their course programs) (see 4.5.3).

The sources of these assessment cognitions included teacher training, books, teaching experience, colleagues (both inside and outside the organization in which the teachers worked), and other personal experiences (see 4.6).

7.2.2 Areas of possible weakness in assessment cognition

Besides mapping the cognitions involved in classroom assessment, this study also found and probed areas where teachers' judgments seemed to be particularly vulnerable to error. One such area was in the interactive assessment of students' receptive skills, especially listening (see 6.3.2). Unlike with speaking and writing, where there was often a product involved, there was often no obvious evidence on which to base assessments of listening and reading. With listening, teachers gained some evidence when a student interacted with the teacher or with other students under the teacher's observation, but this was problematic in that other possible dimensions of the listening ability construct, such as understanding lectures, were excluded. One way that was suggested of bolstering teachers' interactive assessments in the first area was to complement them with formal tests (a tactic also mentioned in Harlen & James, 1997), although teachers were wary of over-testing students and reducing precious teaching time.

Another area of possible weakness was misjudging student activeness in class or lack thereof. With the former, an active student could give off an impression that would lead the teacher to overestimate that student's linguistic abilities. With the latter, there was an inherent ambiguity in student silence: that is, teachers could not know whether a silent student truly did not understand something spoken (or written, for that matter), or understood it but merely did not wish to speak. To deal with this area

of possible weakness, it was suggested that changing student groupings and using a variety of subject texts to balance quiet and active students' classroom interaction performance. Somewhat related to this, a general way of improving teachers' impressions that was mentioned was to try and avoid judging students prematurely, at the start of a class.

A third possible weakness was that teachers based their projections largely on personal experience, rather than on specific knowledge of actual target language use situations (see 6.5.2). Teachers usually drew upon their own academic experiences or upon information incidentally gathered from past students who had gone on to their degree courses, rather than upon systematically-gained knowledge of the language demands of those degree courses. This suggests a possible gap between the expectations that teachers imagined course tutors had and thus applied to student performances on one hand and the actual expectations of those course tutors on the other. Ways of strengthening projections are mentioned in the following sections.

7.2.3 Organizational influences upon assessment cognitions

While it was apparent from the data in Stages 1 and 2 that a teacher's assessment cognition network was very personal and individual to the teacher because it was founded upon personal experience and beliefs about teaching, learning, and assessment, it was also very clear that the organization within which the teacher taught courses exerted a very strong influence upon the network. Exploring this issue, this study found specific ways in which that influence was exerted.

One was that summative assessment tasks set at the organizational level (i.e., by decision of a manager and/or teacher group) were often used in modified form for formative assessment in the classroom (see III under 4.5.1). It can thus be said that in instances when this was done by the teacher, the organization shaped or structured teachers' assessment cognitions via the task, albeit in a weak sense, as the task was still mediated in practice by the teacher and adjusted to fit pedagogical purposes.

Another way the organization exerted influence on assessment cognition was through the syllabus. The degree of that influence seemed to depend on teachers'

subjective perception of pace (see 6.4 and III under 4.5.1); that is, the more a teacher felt the syllabus content was too much to cover within a given time frame, especially in view of the teacher's personal pedagogical approach and ability level of students in that class, the stronger the sense of time pressure. This in turn affected the teacher's assessment practices, as time concerns became a weightier consideration; for example, teachers had to cut short their feedback.

In addition, it could be said that the organization had a strong influence on assessment cognitions in a very basic manner: by setting the class parameters or realities, such as number of students in a class and class time duration and frequency. Considerations of such parameters also figured heavily in teachers' assessment thinking (see II under 4.5.1).

The study also found possible ways in which the organization could strengthen teachers' assessment cognition networks. One was through its formal testing; information from an organization's placement tests could add to teachers' knowledge of students (see 6.3). Another was through conducting studies about students and the linguistic demands they faced on their degree courses (see 6.5.2), thus strengthening the assessment cognition of projection.

Below, I discuss how these findings confirm and extend current knowledge in the area of teachers' assessment cognitions, after a discussion of the study's limitations.

7.3 Limitations of the study

While I believe the findings summarized above are firmly grounded in the data and will prove fairly robust in light of future research, I also recognize that there need to be some caveats in view of the study's limitations. Two such limitations strike me as particularly important.

The first is the problem of segmentation. I chose the video "episodes" for the teachers to comment upon. While, as I have argued in 3.4.4, the episodes I chose did involve assessment, I think I should have incorporated Eraut's (1994) idea of examining a professional's "performance period." This means examining everything

done by the professional during a specified period of time, like from lunch time to tea-break or from start to finish of a task:

What is important ... is that all the tasks or transactions performed during the period are included, even if they have no connection with each other beyond their claims on the performer's attention. Without this requirement, a major feature of some occupations would be excluded, namely the handling of competing demands. (p. 150)

Thus, during one of the stimulated recall sessions, I could have had the Stage 1 teachers try to recall, as much as possible, what they were thinking during a whole class period; while I think a strength of this study is its consideration of the “natural” classroom context, this addition would have allowed me to see assessment cognitions in the wider “flow” of teacher thoughts, lending even more “ecological validity” to my analysis.

Secondly, this study is limited in its contribution to the ambitions of the critical realist paradigm I set forth in 3.2. An ongoing interest of work from this paradigm position is how social structures influence social agents.² While I think this study has succeeded in drawing out some of those influences (e.g., of the institution and the profession), it is limited to what the teachers are aware of; there may be important influences that the teachers were not aware of. One such possible influence is external examination arrangements.³ These did not seem to impinge on teachers’ day-to-day thinking, but CTA, the teacher-manager in Stage 1, was aware that the course syllabi and summative assessments would need to pass external examination [A3:100-106]. In addition, data from an evaluation I conducted for one of my doctoral course assignments, about the effectiveness of a postgraduate course assessment policy, also suggested that the choice of summative assessment method was partly constrained by the fact that it had to withstand external examiner scrutiny. If this is so, then one can see that there is an indirect influence on the classroom—a kind of “washback effect” of the external examination process—because teachers may have students practise a task similar to the summative assessment task. This study was limited to the influences that teachers perceived; further research into the social structures—perceived by teachers or not—that impact classroom assessment

² Of course, it is also interested in how agents influence structures, too.

³ I was alerted to this possibility because the language center in Stage 1 had a visit from external examiners during the term I observed; the teachers expressed some anxiety at the time but did not mention it at all in relation to their classroom work.

remains to be done. In terms of the dimensions I discussed in 2.2.1, this would go beyond looking at the cognitive side of classroom assessment and involve more sociologically-oriented research.

These limitations aside, I do think this study has contributed in a number of ways to knowledge about teachers' assessment thinking. It is to these contributions that I now turn.

7.4 Contribution to knowledge

This study confirms and extends many aspects of what is known about teachers' assessment thinking. In this section, I integrate the findings of my study with current knowledge and discuss the specific contributions that this study makes.

7.4.1 Originality of methodology

Before discussing these specific contributions in detail, I should point out that I believe these contributions arise in large part from the originality of this study's research design; while previous research has provided valuable insights into teachers' assessment thinking, this study has been able to explore some of the "blind spots" in previous studies deriving from their methodology. First, a number of previous studies (Davison, 2004; Orrell, 1995; Reali et al., 2001), usually using think aloud protocols, have looked at what teachers think about as they assess student work. These have shed light on assessment thought processes but largely for summative purposes and detached from the classroom. This study, in contrast, has examined thought processes involved in assessment in the classroom, largely for formative purposes. Second, other previous studies (Bachor & Anderson, 1994; Hall & Harding, 2002; Hall et al., 1997; McCallum et al., 1993; McMillan & Nash, 2000) have looked at classroom assessment practices and the thinking behind such practices, but this has mainly been done through interviewing rather than through classroom observation. This study, on the other hand, employed classroom observation and looked at assessment practices and thinking *in situ* (in Stage 1). Third, the few studies of assessment thinking that have involved classroom observation either only looked at attitudes but not thought processes (Yung, 2001 &

2002) or explored assessment thought processes but only hinted at the wider influences upon such processes (Edelenbos & Kubanek-German, 2004; Mavrommatis, 1997; Rea-Dickins, 2001). This study has explored those wider influences at length, both through the stimulated recall sessions and through the focus groups. Finally, almost all the studies in this area of assessment cognition have looked at primary or secondary education, while this study has studied EAP, which has a very different student composition and policy environment, among other differences. The originality of the research design thus resulted in findings that both supported previous studies and added new insights; these are detailed below.

7.4.2 Teachers' individual assessment practices and cognitions

As the literature review in chapter 2 made clear, previous studies (e.g., Bachor & Anderson, 1994; Davison, 2004; Edelenbos & Kubanek-German, 2004; McCallum et al., 1993; McMillan & Nash, 2000; Reali et al., 2001; Yung, 2001 & 2002) found that teachers had a constellation of assessment practices and accompanying assessment cognitions that were fairly unique to each teacher. This study not only confirmed this but also, through its in-depth examination of two cases, mapped the influences—such as beliefs about teaching, learning, and assessment, and organizational requirements—that shaped and produced that unique constellation, what this study has called the teacher's assessment cognition network. These studies, while affirming the individuality and diversity of teachers' assessment cognitions, also suggested that teachers could be grouped in typologies according to similarities in their assessment practices and cognitions. The mapping done in this study suggests possible sources of these similarities: perhaps teachers having similar assessment practices and cognitions had similar teacher training (thus leading to like beliefs about teaching, learning, and assessment), or participated in common or similar professional communities or educational organizations.

Studies such as Mavrommatis (1997) and Rea-Dickins & Gardner (2000)⁴ have also found possible weaknesses in teachers' assessment practices and cognitions. For example, Rea-Dickins & Gardner found that primary teachers sometimes made recording errors when assessing their EAL students. While this study's data did not

⁴ This article was not included in the literature review but was also based on the larger study from which Rea-Dickins (2001) drew its data.

reveal such obvious errors, it did suggest that EAP teachers had difficulty assessing students' receptive skills, and that their initial judgments of students could be mistaken (see 6.3).

7.4.3 The influence of external factors, particularly the organization

Because many of the studies previously conducted on this topic of teachers' assessment thinking have focused on primary and secondary education, they have framed assessment in relation to national or state assessment policies (e.g., England's National Curriculum in McCallum et al., 1993; British Columbia's assessment policy reforms in Bachor & Anderson, 1994). This is in recognition of the fact that such mandated policies can have an important impact upon teacher thinking. This study, in examining the EAP context, confirms that fact but locates the mandated policies not at a government level but at the organizational level. While there is obviously mediation of government policy at the primary and secondary school level, thus influencing classroom practice (as Hall & Harding's 2002 study of "communities of assessment practice" attempts to describe), there is still a strong connection between government policy and classroom action; however, there are not such strong government policy imperatives bearing upon university language centers teaching EAP, and so the organization does not so much mediate policy as it does set policy in the form of course syllabi and summative assessments. One of this study's contributions to current knowledge therefore is its exploration of what has previously been somewhat ignored in the literature—the influence of the organization upon teachers' assessment cognitions.⁵

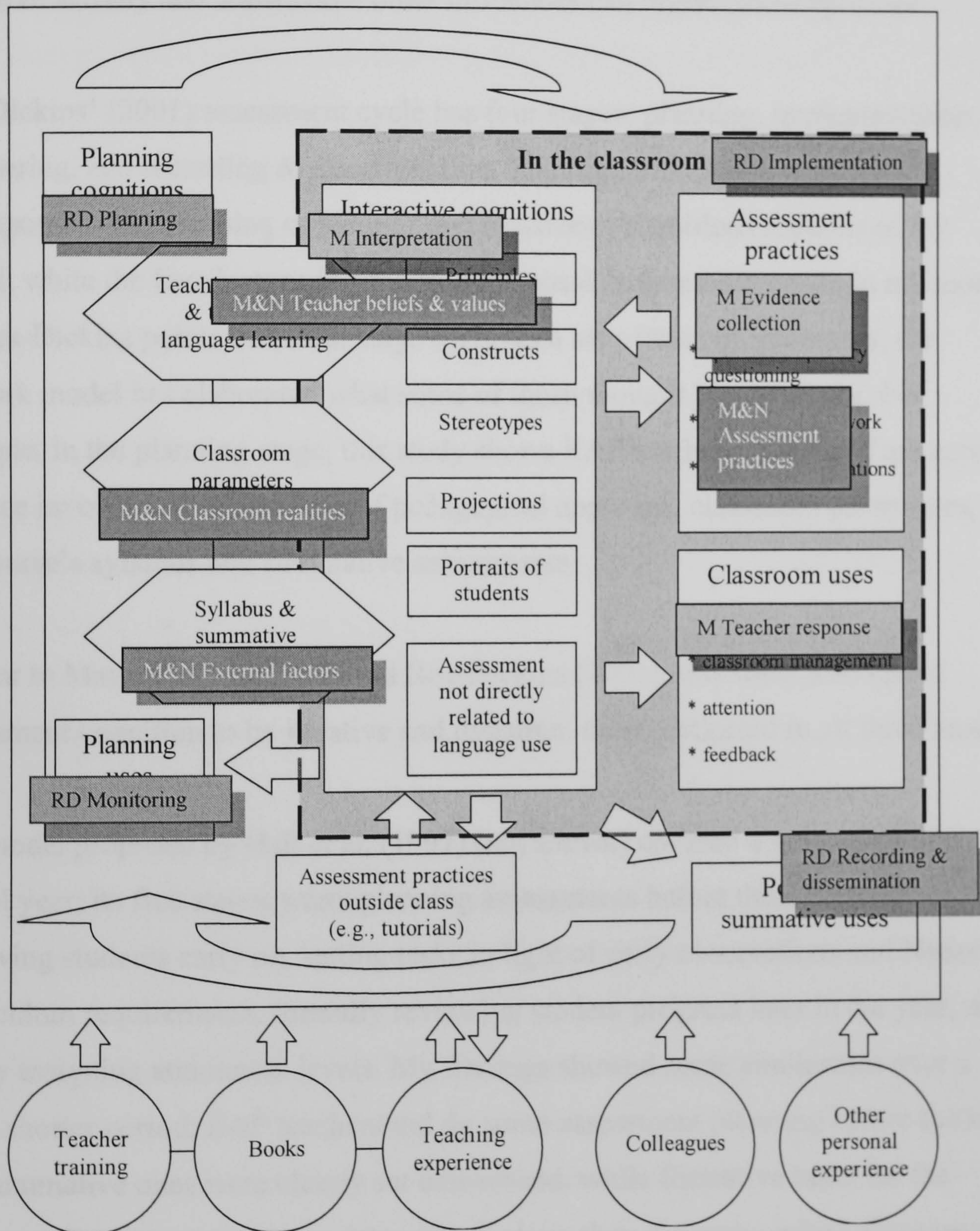
7.4.4 Models of teacher thinking in assessment

The chapter 2 literature review also presented models of assessment thinking based upon previous empirical work (see 2.4.1). Since there had not been any studies of assessment thinking that took a broader definition of assessment and that looked at EAP, as I wanted to, I decided to take a more grounded approach to my analysis, starting off with only a few pre-determined concepts. I think this is why the model of the assessment cognition network I have put forth in Figure 7.1 does not bear a

⁵ This study of organizational influence also suggests that an examination of the influence of the wider institution—the university—is also in order.

surface resemblance to the models reviewed. However, if one looks carefully, one can see strong similarities between my network model and previous models. I illustrate these in Figure 7.2 and then compare my model with the others in turn.

Figure 7.2 Aspects of previous models that are similar to the assessment cognition network model



Key

M: from Mavrommatis (1997)

RD: from Rea-Dickins (2001)

M&N: from McMillan & Nash (2000)

Mavrommatis (1997) characterizes an “assessment episode” as occurring in evidence collection, interpretation, teacher response, and impact on students. His first three steps resemble the assessment practices, interactive cognitions, and uses parts of my model. Besides examining the nature of such assessment episodes in the context of EAP, my model also sets these “assessment episodes” in a larger network of cognitive activity and shows how other influences can impact these episodes.

Rea-Dickins’ (2001) assessment cycle has four stages: planning, implementation, monitoring, and recording & dissemination. The first three stages roughly correspond to the planning cognitions and classroom cognitions sections of my model, while the fourth stage is similar to the possible summative uses in my model. As Rea-Dickins points out, each stage involves a set of thought processes; the network model has elaborated what some of those thought processes are. For example, in the planning stage, this study shows EAP teachers’ choice of assessment practice involved considerations of pedagogical approach, classroom parameters, and the course’s syllabus and summative assessments.

Similar to Mavrommatis (1997) and Rea-Dickins (2001), this study also found assessment cognition to be iterative and dynamic; this is reflected in all three models.

The model proposed by Hall et al. (1997) (not shown) covered a whole primary school year; its five stages were: planning assessments before the school year, observing students early on, setting tasks in light of early observations and National Curriculum requirements, formally reviewing student progress later in the year, and finally assigning attainment levels. My findings showed some similarities over a much shorter period; EAP teachers did do some assessment planning before the term (the summative ones were clearly set beforehand, while formative tasks for the classroom were sometimes pre-determined, also), they did make initial diagnoses of students’ abilities, and they did adjust their assessments as the term went on. The informal assessments then could play a role in summative assessment: for example, a student’s grade on a group discussion could be raised if it seemed out of line with his or her teacher’s knowledge of the student from class.

McMillan & Nash's (2000) model posited that teacher beliefs and values were constantly in tension with classroom realities and external factors: in the midst of dealing with this tension, teachers made decisions about assessment and grading practices. Although this study took a broader definition of assessment, it found much in common with McMillan & Nash's study:

- teachers' beliefs about teaching and learning exerted considerable influence in planning, while teachers' assessment principles mediated interactive assessment;
- classroom realities (what I've called parameters) and external factors (largely organizational in nature, in the case of EAP) were also significant influences;
- teachers did face tension as they managed these sometimes conflicting influences;
- this all impacted upon assessment practices.

As the above discussion shows, my model overlaps and confirms many aspects of previous models of assessment thinking. In addition, this study found other cognitions (like stereotypes) and possible sources of assessment cognitions (e.g., teacher training, books, and colleagues).

In summary, this study's contribution consists of its empirical support for the conclusions of previous research and its addition of new findings to the current body of knowledge about teacher cognition as it relates to assessment. In the final section, I make several recommendations based on the findings of this study.

7.5 Recommendations

While I think the findings of this study are of intrinsic interest, I also believe they have important wider implications. Below, I make recommendations for research, policy, and professional development in light of what this study has revealed.

7.5.1 Research

On one hand, it was very obvious from this study that the organization had important influences upon teachers' assessment cognitions. On the other hand, there has been a dearth of research into this dimension or level. Therefore, one recommendation is

that researchers examine the nature and methods of organizational influence upon classroom assessment. Ideally, just as Broadfoot (1996) and other studies have considered the key functions that summative assessment serves in a society, future research would consider the roles that assessment (perhaps formative but more likely summative) serves within organizations and the organizational imperatives that shape decision-making related to assessment. Another possible area of research at the organizational level is what can be called “micro-washback,” the influence of organizationally mandated summative assessments upon teaching and learning. Research at the organizational level could also examine, as Hall & Harding (2002) do, the organization as an assessment community. One intriguing (and unintended) finding from this study was that teachers within a language center used a common shorthand (in two of the programs studied, IELTS levels were the basis) to refer to students. I believe this kind of phenomenon, and the possible ramifications it has for classroom assessment, require investigation. There are of course challenges in any undertaking of the kind recommended here, one prominent one being finding appropriate conceptual tools—perhaps from organizational theory—that can help researchers approach and analyze this topic. However, overcoming these challenges would provide a richer and more complete picture of the context in which teachers think about and conduct classroom assessment.

I also recommend examining how particular teaching methods afford particular assessment practices and thus particular ways of knowing students, or conversely, how particular assessment practices may afford particular teaching methods (as the findings from Black, Harrison, Lee, Marshall, & Wiliam (2003) would indicate). While I wholeheartedly agree with the idea that it is teachers, not methods per se, that act in the classroom to help students learn, it still seems the teachers observed in this study managed their classrooms and lessons with at least general beliefs and principles of particular teaching methodologies in mind (e.g., present-practice-produce or discovery learning), and this inherently involved particular assessment practices (e.g., calling upon students or going around to observe group work). As far as I am aware, this link between teaching method and assessment practices has not been explicitly researched. But as the idea that the processes of teaching and assessment are inextricably bound together gains greater currency, the need to

understand this link becomes greater. Greater awareness of this link through research can also inform teacher training (see 7.5.3).

The final recommendation I have in this area is simply that more research needs to be done into teacher thinking in classroom assessment, but that research needs to be characterized by a respect for its highly situated nature. Brookhart (2004, p. 447) has pointed out in her survey of classroom assessment research that many studies have applied criteria from educational measurement to aspects of classroom assessment, including teachers' assessment knowledge; in all of these, "someone or something is found wanting." However, I think such a prescriptive approach or deficit view is effectively blind not only to the great deal of cognitive activities and resources that teachers bring to classroom assessment, but also to the diverse and often conflicting demands that teachers must mediate as they conduct assessment in the classroom. I do not wish to privilege teacher thinking beyond constructive critique, but I do believe research in this area needs to recognize that teacher thinking in assessment involves far more than just appraising a student, as this study has shown.

Knowledge gained from the recommendations for research mentioned here can feed into policymaking and professional development. It is to these two areas that I now turn.

7.5.2 Policy

Based on the findings of this study, two recommendations for policymakers at various levels can be made. One general recommendation is that policymakers must account for the components of teachers' assessment cognition networks whenever assessment reform is desired; such reforms at whatever level—national, state, district, or organization—will need to either align with or influence such things as pedagogical approaches and beliefs, classroom parameters like class size and scheduling, and implicit constructs if they are to have any chance of successful implementation.

The second recommendation is that organizational policymakers consider ways in which the organization can leverage policies so as to strengthen teachers' assessment

cognition networks. This is especially relevant to language centers providing EAP support, as they will presumably be evaluated by other university departments (at the same or other institutions) on the basis of the quality or accuracy of the centers' summative appraisals of students going through its programs. Thus, it is in the organization's best interest to support teachers' assessment thinking and practices. This study's findings suggest at least two possible means. One would be to review the syllabus (e.g., is there too much content to cover too quickly?) and summative assessments (e.g., are there any "micro-washback" effects upon the teaching and learning that occurs in our classes?); one of this study's most apparent findings was that these "policies" had direct and influential impacts upon classroom assessment. Another means would be to implement policies that would support teacher research to improve parts of the cognition network, especially stereotypes and projections. In regards to the latter, I have a specific suggestion—as projections were examined in some depth over the two stages of this study—that can serve as an example of the kinds of organizationally supported research projects I am envisioning. This study found that EAP teachers judged their students' performance by projecting it into the target language use situation (in this case, academic situations students may face on their degree programs) and imagining how tutors or classmates would assess it. The teachers' knowledge of the situations and how performances would be judged was based largely on personal experiences with academia and partly on discussions with students on their degree courses. It seems to me that such knowledge can be put on a stronger foundation, through—aptly enough—drawing on an idea in Sadler's (1989) article on formative assessment. Sadler proposes having teachers use exemplars with their learners and discuss what characteristics of these exemplar performances constitute differing levels of quality; in this way, learners will have a clearer idea of how work will be assessed. It is also important, Sadler adds, that a number of exemplars be given so that students will

learn that there are different ways in which work of a particular quality can find expression. There is often a wide variety of objects within the same genre which are regarded as excellent. Unless students come to this understanding, and learn how to abstract the qualities which run across cases with different surface features but which are judged equivalent, they can hardly be said to appreciate the concept of quality at all. (p. 128)

EAP teachers may be able to have a much clearer idea of how degree program tutors judge student performance if they can take the role of learners in Sadler's proposal. That is, by having course tutors make their implicit assessment criteria more explicit

through the use of exemplars, EAP teachers can abstract and then “internalize,” so to speak, general quality criteria and employ them as they assess student work.⁶ In testing terms (see Douglas, 2000), this would go beyond simulating the linguistic tasks students will face in their programs (situational authenticity)—usually based on a needs analysis—and attempt to simulate more closely the ways in which those tasks will be assessed (part of interactional authenticity). An organizational research project like this would involve the kind of collaboration that Douglas (2000), writing about the construction of assessments of languages for specific purposes, recommends between testing experts and subject specialists (in this case, between teachers and degree program tutors).

7.5.3 *Professional development*

As with policy, I have two recommendations for encouraging professional development in the area of classroom assessment based upon these findings. The first relates to language teacher training: rethink and renovate the language assessment curriculum in teacher training courses. This is based not so much on the presence of a particular finding in my study but on its absence: in both stages of my study, participants hardly mentioned teacher training in assessment; when teacher training was mentioned, it was in relation to their teaching methods and practices rather than their assessment practices per se.⁷ This indicates a disparity between what teachers-in-training are taught and what teachers-at-work actually deal with on a regular basis; this disparity is also supported by personal experience and research (see 1.3). I suggest at least adding topics like 1) the relationship between teaching methods and assessment practices (especially teacher questioning practices), 2) the contextual demands (like organizational policies and classroom parameters) that affect assessment, and 3) the cognitions employed in interactive assessment and how to improve them. These additions would not only increase the “assessment awareness” of teachers-in-training (see next recommendation) but also place the assessment principles and techniques that seem to currently comprise the content of

⁶ In a similar approach, some EAL/ESL assessment frameworks like the ESL Bandscales or NLLIA Bandscales provide exemplars to illustrate its levels (Scott & Erduran, 2004).

⁷ Of course I have argued in this study that teaching practices implicate certain assessment practices, so in this sense teachers were also trained in assessment, but I do not think the training was framed in these terms, nor were the links explicitly discussed.

most assessment training courses within the broader classroom assessment “landscape” that teachers will have to navigate.

My last recommendation is for individual or groups of teachers to consider: it is that teachers use the findings of this study to become more aware of their own assessment cognition networks, the significant influences upon those networks, and the possible weaknesses in those networks that threaten the quality of teachers’ assessments of students. I believe the model I have proposed can serve as a kind of scaffolding tool for teachers, particularly of EAP; the model—and this study’s findings more generally—can lend some structure and focus as teachers reflect on their own assessment practices and thinking. Teachers would also benefit from seeing how they and the organization they work in “co-construct” the judgments that they make about students, which this study has outlined.

7.6 Concluding remarks

As I mentioned in the introductory chapter, I explained how I wanted to conduct this study largely because of my own experiences of dilemmas and mistakes in assessing my students. When I began this study, I think I had hoped to gain some special insight into how to resolve all those dilemmas and how to correct all my mistakes in assessing students. Having completed this study, I think I can say that I have gained special insight, but not the kind that I had expected—I now see why I cannot resolve all those conflicts or avoid all the mistakes: they are inherent to the work of teaching because of, as this study has found, how classroom assessment is situated at the intersection of so many demands and considerations, and how personal experiences and (always limited) knowledge shape teachers’ judgments of students. However, this study has also shown me ways forward, so that while those conflicts and errors in assessment will never be eliminated, they can be managed and reduced. It is my hope that other educators, especially teachers, can use the findings of this study to make themselves more aware of the nature of the various difficulties and problems that teachers face when they assess students, and in so doing work towards dealing with those challenges in ways that will improve teachers’ assessment thinking and practices and ultimately help the students they serve reach their teaching and learning goals.

Bibliography

- Airasian, P. W. (1997). *Classroom assessment* (3rd ed.). London: McGraw-Hill.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Arkoudis, S., & O'Loughlin, K. (2004). Tensions between validity and outcomes: Teacher assessment of written work of recently arrived immigrant ESL students. *Language Testing*, 21(3), 284-304.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachor, D. G., & Anderson, J. O. (1994). Elementary teachers' assessment practices as observed in the province of British Columbia. *Assessment in Education*, 1(1), 63-93.
- Bailey, K. M. (1996a). The best laid plans: Teachers' in-class decisions to depart from their lesson plans. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom* (pp. 15-40). Cambridge: Cambridge University Press.
- Bailey, K. M. (1996b). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines* (2nd ed.). Buckingham: Open University Press.
- Bhaskar, R. (2000). *From east to west: Odyssey of a soul*. London: Routledge.
- Bickel, W. C., & Hatrup, R. (1995). Teachers and researchers in collaboration. *American Educational Research Journal*, 32(1), 35-65.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research for education: An introduction to theory and methods*. London: Allyn & Bacon.
- Bond, T. (2000). *Standards and ethics for counselling in action* (2nd ed.). London: Sage.
- Borg, S. (1998). Teachers' pedagogical systems and grammar teaching: A qualitative study. *TESOL Quarterly*, 32(1), 9-38.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what teachers think, know, believe, and do. *Language Teaching*, 36, 81-109.

- Breen, M. P., Hird, B., Milton, M., Oliver, R., & Thwaite, A. (2001). Making sense of language teaching: Teachers' principles and classroom practices. *Applied Linguistics*, 22(4), 470-501.
- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing*, 15(1), 45-85.
- Broadfoot, P. (1996). *Education, assessment and society: A sociological analysis*. Buckingham: Open University Press.
- Broadfoot, P., & Pollard, A. (2000). The changing discourses of assessment policy: The case of English primary education. In A. Filer (Ed.), *Assessment: Social practice and social product* (pp. 11-26). London: Routledge Falmer.
- Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education*, 8(2), 153-169.
- Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429-458.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
- Bryman, A., & Burgess, R. G. (1994). Developments in qualitative data analysis: An introduction. In A. Bryman & R. G. Burgess (Eds.), *Analyzing qualitative data* (pp. 1-17). London: Routledge.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Calderhead, J. (1981). Stimulated recall: A method for research on teaching. *British Journal of Educational Psychology*, 51, 211-217.
- Calderhead, J. (1984). *Teachers' classroom decision-making*. London: Holt, Rinehart & Winston.
- Cheng, L. (2000). Washback or backwash: A review of the impact of testing on teaching and learning. (ERIC Document Reproduction Service No. ED 442 280)
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389.
- Clandinin, D. J. (1986). *Classroom practice: Teacher images in action*. London: Falmer.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). London: Collier Macmillan.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.

- Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. London: Sage.
- Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: Specific purposes or general purposes? *Language Testing*, 18(2), 207-224.
- Cumming, A. (2003). Experienced ESL/EFL writing instructors' conceptualizations of their teaching: Curriculum options and implications. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 71-92). Cambridge: Cambridge University Press.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334.
- Deem, R. (2004). The knowledge worker, the manager-academic and the contemporary UK university: New and old forms of public management? *Financial Accountability and Management*, 20(2), 107-128.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of 'diagnostic competence'. *Language Testing*, 21(3), 258-282.
- Elbaz, F. (1983). *Teacher thinking: A study of practical knowledge*. London: Croom Helm.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (1998). Teaching and research: Options in grammar teaching. *TESOL Quarterly*, 32(1), 39-60.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Eraut, M. (1994). *Developing professional knowledge and competence*. London: Falmer Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). London: The MIT Press.
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, 38(1), 47-65.
- Filer, A. (Ed.). (2000). *Assessment: Social practice and social product*. London: Routledge Falmer.
- Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 645-672). Thousand Oaks, CA: Sage.

- Freeman, D. (2002). The hidden side of the work: Teacher knowledge and learning to teach. *Language Teaching*, 35, 1-13.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Guba, E. C. (1990). The alternative paradigm dialog. In E. C. Guba (Ed.), *The paradigm dialog* (pp. 17-27). London: Sage.
- Hall, K., & Harding, A. (2002). Level descriptions and teacher assessment in England: Towards a community of assessment practice. *Educational Research*, 44(1), 1-15.
- Hall, K., Webber, B., Varley, S., Young, V., & Dorman, P. (1997). A study of teacher assessment at Key Stage One. *Cambridge Journal of Education*, 27(1), 107-122.
- Harlen, W., & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning, *Research Evidence in Education Library (REEL)*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, 4(3), 365-379.
- Hartwig, M. (2001). New Left, New Age, New Paradigm? Roy Bhaskar's *From East to West*. *Journal for the Theory of Social Behaviour*, 31(2), 139-165.
- Huberman, A. M. (1990). Linkage between researchers and practitioners: A qualitative study. *American Educational Research Journal*, 27(2), 363-391.
- Huerta-Macías, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5, 8-11.
- Jackson, P. W. (1968). *Life in classrooms*. New York: Holt, Rinehart, & Winston.
- Jordan, R. R. (2002). The growth of EAP in Britain. *Journal of English for Academic Purposes*, 1, 69-78.
- Kelle, U. (Ed.). (1995). *Computer-aided qualitative data analysis: Theory, methods and practice*. London: Sage.
- Kennedy, M. (1997). The connection between research and practice. *Educational Researcher*, 26(7), 4-12.
- Kimmel, A. J. (1988). *Ethics and values in applied social research*. London: Sage.

- Kramsch, C. (1995). The applied linguist and the foreign language teacher: Can they talk to each other? In G. Cook & B. Seidlhofer (Eds.), *Principle & practice in applied linguistics*. Oxford: Oxford University Press.
- Krueger, R. A. (1994). *Focus groups: A practical guide for applied research* (2nd ed.). London: Sage.
- Kuckartz, U. (2001). *MAX qualitative data analysis introduction*. Berlin: VERBI Software.
- Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. London: Sage.
- Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly*, 1(1), 19-41.
- Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing*, 21(3), 334-358.
- Leung, C., & Teasdale, A. (1997). What do teachers mean by speaking and listening? A contextualised study of assessment in multilingual classrooms in the english national curriculum. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (Eds.), *Current developments and alternatives in language assessment: Proceedings of LTRC '96* (pp. 291-324). Jyväskylä, Finland: University of Jyväskylä.
- Marshall, C., & Rossman, G. B. (1995). *Designing qualitative research* (2nd ed.). London: Sage.
- Mavrommatis, Y. (1997). Understanding assessment in the classroom: Phases of the assessment process--the assessment episode. *Assessment in Education*, 4(3), 381-399.
- McCallum, B., McAlister, S., Brown, M., & Gipps, C. (1993). Teacher assessment at Key Stage One. *Research Papers in Education*, 8(3), 305-327.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 20(4), 34-43.
- McMillan, J. H., & Nash, S. (2000). Teacher classroom assessment and grading practices decision making, *Annual Meeting of the National Council on Measurement in Education*. New Orleans, Louisiana.
- McNamara, T. (1996). *Measuring second language performance*. London: Longman.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333-349.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). London: Sage.
- Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). London: Sage.
- Morgan, D. L. (2002). Focus group interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context & method* (pp. 141-160). London: Sage.
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13-25.
- Myers, G., & Macnaghten, P. (1999). Can focus groups be analysed as talk? In R. S. Barbour & J. Kitzinger (Eds.), *Developing focus group research* (pp. 173-185). London: Sage.
- Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Orrell, J. (1995). *Is past wisdom enough to ensure future quality in higher education?* Paper presented at the Australian Association for Research in Education Conference: Future Directions, Townsville, QLD Australia.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332.
- Phillips, D. C. (1990). Postpositivistic science: Myths and realities. In E. C. Guba (Ed.), *The paradigm dialog* (pp. 31-45). London: Sage.
- Popham, W. J. (2002). *Classroom assessment: What teachers need to know*. London: Allyn & Bacon.
- Porter, S. (2002). Critical realist ethnography. In T. May (Ed.), *Qualitative research in action* (pp. 53-72). London: Sage.
- Punch, K. F. (1998). *Introduction to social research: Quantitative and qualitative approaches*. London: Sage.
- Raveaud, M. (2004). Assessment in French and English infant schools: Assessing the work, the child or the culture? *Assessment in Education*, 11(2), 193-211.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Rea-Dickins, P. (2003). Framing assessment: Breaking the mould. Plenary paper. *TESOL Symposium on ESL EFL Standards for Young Learners* (pp. 17-29). University of Rome. September. Alexandria, Virginia: TESOL, Inc.
- Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing*, 17(2), 215-243.

- Reali, A. M. d. M. R., Reyes, C. R., Martucci, E. M., Mizukami, M. d. G. N., Lima, E. F. d., Tancredi, R. M. S. P., & Mello, R. R. (2001). How teachers assess and attribute grades: Report on a teaching and learning experience. *Annual Meeting of the International Association for Educational Assessment*. Rio de Janeiro, Brazil.
- Richards, J. C. (1998). *Beyond training: Perspectives on language teacher education*. Cambridge: Cambridge University Press.
- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (2002). Influences on student cognitions about evaluation. *Assessment in Education*, 9(1), 81-95.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Salomon, G. (Ed.). (1993). *Distributed cognitions: Psychological and educational considerations*. Cambridge: Cambridge University Press.
- Samuelowicz, K., & Bain, J. D. (2002). Identifying academics' orientations to assessment practice. *Higher Education*, 43, 173-201.
- Schaeffer, F. (1968). *Escape from reason*. London: Inter-Varsity Fellowship.
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition*. Rowley, MA: Newbury House.
- Scott, C. (2005). Washback in the UK primary context with EAL learners: Exploratory case studies. Unpublished doctoral dissertation, University of Bristol.
- Scott, C., & Erduran, S. (2004). Learning from international frameworks for assessment: EAL descriptors in Australia and the USA. *Language Testing*, 21(3), 409-431.
- Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgements, decisions, and behavior. *Review of Educational Research*, 51(4), 455-498.
- Silverman, D. (2001). *Interpreting qualitative data: Methods for analyzing talk, text and interaction* (2nd ed.). London: Sage.
- Stake, R. E. (1995). *The art of case study research*. London: Sage.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, New York: State University of New York Press.
- Stough, L. M., & Emmer, E. T. (1998). Teachers' emotions and test feedback. *Qualitative Studies in Education*, 11(2), 341-361.

- Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, 17(2), 163-184.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham: Open University Press.
- Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22(4), 389-404.
- Vaughn, S., Schumm, J. S., & Sinagub, J. (1996). *Focus group interviews in education and psychology*. London: Sage.
- Wall, D., & Alderson, J. C. (1996). Examining washback: The Sri Lankan impact study. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 194-221). Clevedon: Multilingual matters.
- Wolcott, H. F. (1990). *Writing up qualitative research*. Newbury Park: Sage.
- Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making and classroom practice*. Cambridge: Cambridge University Press.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). London: Sage.
- Yung, B. H. W. (2001). Examiner, policeman or students' companion: Teachers' perceptions of their role in an assessment reform. *Educational Review*, 53(3), 251-260.
- Yung, B. H. W. (2002). Same assessment, different practice: Professional consciousness as a determinant of teachers' practice in a school-based assessment scheme. *Assessment in Education*, 9(1), 97-117.

Appendix 1 Davison (2004) cline of assessment orientations

Davison (2004, p. 325) proposes that teachers can be located on a spectrum of assessment orientations (see 2.4.3 for further discussion).

	Assessor as technician	Assessor as the interpreter of the law	Assessor as the principled yet pragmatic professional	Assessor as the arbiter of 'community' values	Assessor as God
View of the assessment task	Criterion-bound e.g., It's a checklist ... (A6)	Criteria-based e.g., It's just like a driving test (A10)	Criteria-referenced, but localized accommodations	Community-referenced e.g., It has to be like on the exam (HK1)	Community-bound e.g., It's hard to specify ... (HK3)
View of the assessment process	Mechanistic, procedural, automatic, technical, seemingly universalized e.g., I just follow the criteria (A9); I ticked the boxes (A6); The criteria are just there, so it's really easy (A6)	De-personalized, explicit, codified, legalistic, culturally detached. e.g., I have to be legalistic ... (A6); I would like to give a higher grade but I can't because of the criteria (A10)	Principled, explicit but interpretative, attuned to local cultures/norms/expectations e.g., It's very complex and ultimately you have to give more weight to one thing than another, it comes down to professional judgement (A4)	Personalized, implicit, highly impressionistic culturally – bound e.g., Can this child cop?e (A6); You've saying this child is not tertiary material (A5)	Personalized, intuitive, beyond analysis e.g., You just know ... (HK5); She's just got it ... (A3)
View of the assessment product	Text-focused	Text-focused, but awareness of student	Text and student focused	Student-focused	Student-focused
View of inconsistencies	Seemingly unaffected by inconsistencies	Inconsistencies a problem, threat to reliability e.g., I worry when I make judgment ... am I interpreting the criteria correctly? (A2)	Inconsistencies inevitable, cannot necessarily be resolved satisfactorily, teachers need to rely on professional judgement e.g., I have to juggle things, weight them up in my own mind and think what the alternatives are (A1)	Inconsistencies a problem, threat to validity, assessor training needs to be improved e.g., I think would my colleagues accept this as an A7 (HK2)	Seemingly unaffected by inconsistencies System not open to scrutiny, not accountable, operated by the 'chosen' few.
View of assessor needs e.g., for support/training	Need better assessment criteria	Need better assessor training (to interpret criteria)	Need more time for moderation and professional dialogue (to make basis of judgments more explicit)	Need 'better' assessors (to uphold standards)	

Key : A = Melbourne teacher; HK = Hong Kong teachers.

Appendix 2 Example Stage 1 field notes

This set of observation notes from Stage 1 was chosen because of its relative legibility and its illustration of how I took lower-inference descriptive notes on the left and made higher-inference comments to myself on the right (see I under 3.4.3). For the sake of anonymity, names and initials have been pasted over in several places.

OBSERVATION

CTB 2.24.04 W

5:08 B introduces word-association

B comes over to me to explain his plan
SS are still working.

B says "that's good" to SS

← oops; I was looking more at sheet than observing!

← why?

5:13

SJ
SN
SD
SS SA SY

5:16 B is going around, listening & commenting.
SS are talking

Is assessment going on here?
How do I analyze this?
Observation tied to activity, isn't it?

5:25 B is moving students along, checking, etc.

B gives feedback B corrects SY
B discusses w/ SJ & SN

"correctness"

← feedback

B seems to give students a lot of time/space to talk with each other; his facilitative role can be seen as, really, a "situated teaching" role. The

5:35 B is talking w/ SD & SY

SJ & SN seem to be discussing something task-related; so do SA & SN I think.

5:39 B gets P & N to talk w/ A & S about development as D & Y talk; later, B talks about buzz words and all S listen.

A discusses Attain position (from task sheet) and B corrects "ingenious"

← A mistake, I think. I think SA is trying to say "disingenious."

(I'd imagine he is working on interest vs. syllabus per se.)

B & SS discuss the point regarding development.
B asks them to talk about attribution
B tells me ~~the~~ change of plan.

5:50 B asks A, "what do you think?"
They discuss misrepresentation.
Other pairs are talking or working

} Assessing?

← I have to remember to note other groups, although my focus is on the teacher.

6:04 B asks A about "the scientific method"
Let the class SN

← why?

B continues w/ all SS listening

B focuses on the attribution

Some S. seem to think it's misrepresentation while B says it's off.

← That's my understanding; I might have misheard.

6:18 B goes on to #3

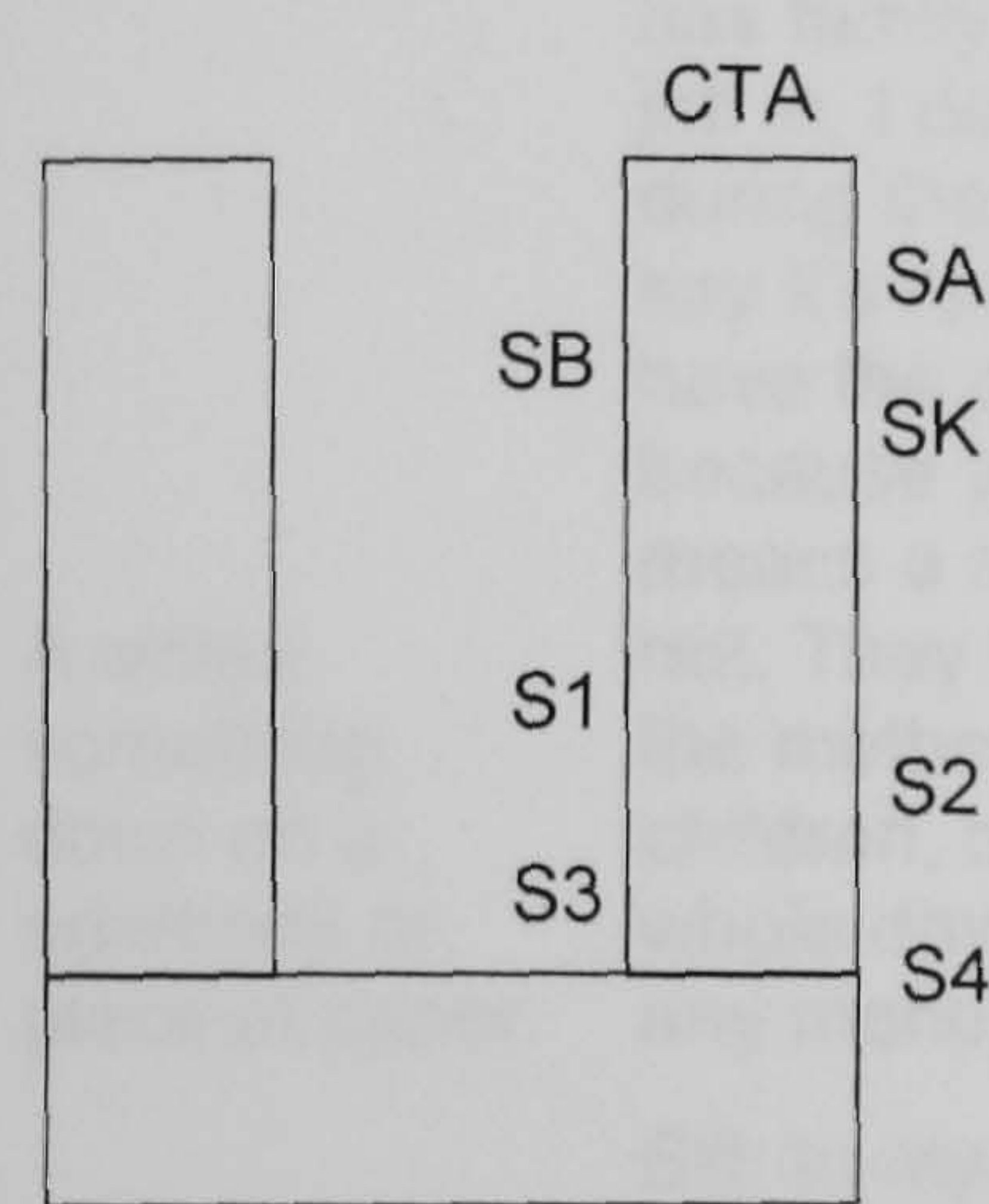
6:20 B hands out task on summarizing
SS (mainly A) discuss.

It seems ~~A~~ ^(P&N) seem to talk quite a bit & class reminds me of my ESL days at OSU!

Appendix 3 Classroom action and teacher recall

This excerpt from Stage 1 illustrates the classroom action that was being shown on the videotape (on the left) alongside the teacher’s verbal report of what he was thinking at the time (on the right) (see also III under 3.4.3 for further discussion). A is the teacher CTA, S and a letter represents students, and M stands for the researcher.

This excerpt is taken from CTA’s EAP class on Friday 6 February 2004. Fridays were set aside for listening and speaking; in this session, the first half hour was concerned with how to argue in a seminar, and students listened to and evaluated two audiotaped speakers about the pros and cons of robots. The next half hour involved students in two groups discussing amongst themselves a “Political Compass” survey that asked students to agree or disagree with controversial statements, thus ostensibly showing them where they fell on the political spectrum. In the last half hour, the two groups held their discussions again, this time taking turns being observed by the teacher and other group. CTA gave feedback to both groups at the end of class.



Below, the first group is about to hold its discussion while being observed by CTA and the other group. The controversial statement the group members are debating is “Those who are able to work, and refuse the opportunity, should not expect society’s support.” The classroom arrangement is diagrammed on the left; CTA is observing group 1 members SB, SA, and SK; S1-4 are group 2 members, who are also observing. In the transcripts below, *** refers to unclear speech, bracketed text with a question mark refer to my guesses at what was said, bracketed text alone is either a paraphrase or a note, and M stands for this researcher. The recall shown here was typical in that the teacher’s recall of his interactive thinking often led to a discussion of wider related topics; for example, the recall of CTA almost giving an answer to SK led to a discussion of the use of summative assessment tasks for formative in-class purposes.

A’s speech & action	Students’ speech & action
<p>A tells group 1 “The floor is yours.”</p> <p>A sits down and looks on at group 1. There is a pause and then SA begins.</p>	

Teacher’s recall
<p>A: That’s the best lesson *** this class so far.</p> <p>M: Yeah, they seemed quite into it.</p> <p>A: They did, didn’t they? Thanks to B; it was B’s choice of lesson. He designed that lesson, he devised that lesson, and it was certainly... the interest was there to get them talking, so I had my data to give them feedback on.</p> <p>[M replays a segment from 2.06 where the students are in groups discussing and A is listening and then giving feedback. The segment runs from 15:03 on the tape to 15:22.]</p>

A's speech & action	Students' speech & action	Teacher's recall
A writes something down on a notebook or piece of paper.	<p>SA: I disagree with this point, because I think it's more comfortable to say to somebody who is really able to work or not, because in our society we have [low unemployment?] *** There is not enough space for everybody. I think society is responsible for everybody.</p> <p>SK: Here it is said if they are able to work but refuse the opportunity means they had the opportunity to work but refuses it. They're not talking about there is not opportunity.</p> <p>SA: *** For example, it could be someone who's living in Bristol and has family here, and they offer him a job in, I don't know, Birmingham during the night. Yeah, you can never say it's really easy to decide if you have the possibility or you haven't, because you need to define what means a real possibility and what's not. They have a situation in America, the mothers have a lot of upset children, but they have to work the whole day; otherwise they won't get any money.</p>	<p>A: [at around 15:05] I'm making notes about lexis.</p> <p>M: Let me pause it.</p> <p>A: OK. I'm making notes on lexis there, and I'm thinking about pronunciation.</p>
A writes something down.	<p>SB: In my opinion, I agree with this point. If the people are able to work, I think the state shouldn't support them. On the other hand, if they haven't got responsibility, if they are disabled people, I think they should have some support from society or government. But generally I agree this point. ...</p>	<p>A: [replay until 15:06] I was noticing her [SB's] use of "on the other hand", nice signpost language, I'd made a note to praise her about that, which I did subsequently.</p>
A writes something down.	<p>SK: I also agree with that point, because from my experience... maybe it depends on the country; I didn't know about the situation in America, but from my experience</p>	
A writes something down.	<p>from Czech Republic, there are a lot of—great majority of the people are gypsy and they don't want to work. They just know that the society and the state will support them. There are places to work, but they just don't want to. They just...</p>	<p>A: [replay until 15:07] [unclear on tape and untranscribed; A talks about noticing SK's fluency, accuracy, and guessing she had public speaking experience.]</p>
	<p>SB: ***</p> <p>SK: Yeah. Yeah. So it's what I see in life. So I agree with that, that they are able to work, they have the opportunity, but they don't do that.</p> <p>SA: You mean they don't work at all and get support from the government.</p>	

A's speech & action	Students' speech & action
	<p>SK: Yeah, mmm. And you can see that, for example, in the city where I live, there are some ... um [glances at CTA] quarter? A part of the city where you got just *** [glances at other group] in the city we have several districts where the gypsy live, and so a normal ordinary day when you go to that district you see a lot of people who are just walking around or drinking beer in bars and they do not work.</p>

Teacher's recall
<p>A: [replay until 15:08] She found the word herself. She used 'quarter' first of all, which is not appropriate in the context, and then she found the word "district." I thought to myself "I'm glad I didn't interrupt her." She was able to correct herself and to use a more appropriate word. So sometimes it's good to keep quiet.</p> <p>M: You chose not to interrupt her because...</p> <p>A: First of all, I wasn't quite sure which word she was searching for, which word she was hunting for. And then also the word she used was good enough, but then she actually found a more appropriate word when she repeated herself.</p> <p>M: If it had been a different word for example, that perhaps would not have been so appropriate, would you have interrupted?</p> <p>A: Yes.</p> <p>M: Even though it's a practice for the assessment, or maybe I'm framing that wrong, because this is partly a run-through of the assessment at the end.</p> <p>A: It's dual purpose. It's primarily to make sure they can function, they can perform well in seminars this week and next week. It's an in-session course, after all, and I've got a responsibility to the department to make sure that students can use English effectively in seminars. Also, because it's also a credit-bearing course and it is assessed, and one of the forms of assessment will be this very mode, they need practice, they need experience in the mode of assessment. So dual purpose.</p> <p>M: Because it's in-class and because of the situation, you're quite willing to give information or interrupt or give feedback in terms of words or whatever, in that kind of situation?</p> <p>A: In that kind of situation. *** I save it until the end, but she was looking for help. If it's a whole-class discussion, I'll interrupt there and then, because it's off-putting; you can't chair a discussion and make notes at the same time. But if I'm purely an observer, as I was there, I'll rely on my notebook and I'll give feedback. I won't undertake any "hot" correction; I'll save feedback until, as I did, until the end of each stage. And that went well today. In that case, she was looking at me for help. I was about to help her but she corrected herself. And it wasn't a gross error, either; it didn't impede understanding.</p>

Appendix 4 Example contact summary form

This is an example contact summary form that helped me organize my data collection and aid my analysis (see 3.5.2). The headings were saved as a template; after each classroom observation in Stage 1, I created a copy of the template, typed in the information after each heading, and saved it as a summary of that class. I then printed it out and stapled it to the front of my notes for that class and any handouts that the teacher had given me and his students.

CTB	Wednesday 4 Feb 04	Summarized 2.05.04	Room 1.8
-----	--------------------	--------------------	----------

Summary of observation

Reading & writing class. The tasks are centered on people's different definitions of words, and representation of others' work.

Impressions:

B follows his "style" of introducing tasks, having students work on them in groups or pairs, going around to listen, ask questions, etc. and then discussing together again.

Students seem fairly engaged, although SN seems strong enough that he finishes earlier than most (that's my assessment), and SA seems to talk quite a bit and asks questions and such in class.

Issues to reflect on:

What is going on in the moments when B is going to each pair? (This is discussed in the I&R.)

Summary of interview after class

We went over a range of issues again: difficulty of amount of syllabus material vs. time constraints, exploring his cognition during the "walking around bit", issues dealing with syllabus and assessment, difficulty of teaching pathway-course (lower level) students who don't respond so much, more about his cognition during online "walking around".

Impressions:

I think I tried to control the talk more, although it's hard because the stuff B talks about is quite fascinating. Of course, maybe he is filling in silence with his extemporaneous thoughts, and maybe I need to be more directive.

There was some interesting stuff we talked about after the tape was done, like managerialism in the university, and rating scales. B gave me stuff he did on rating criteria for his own courses.

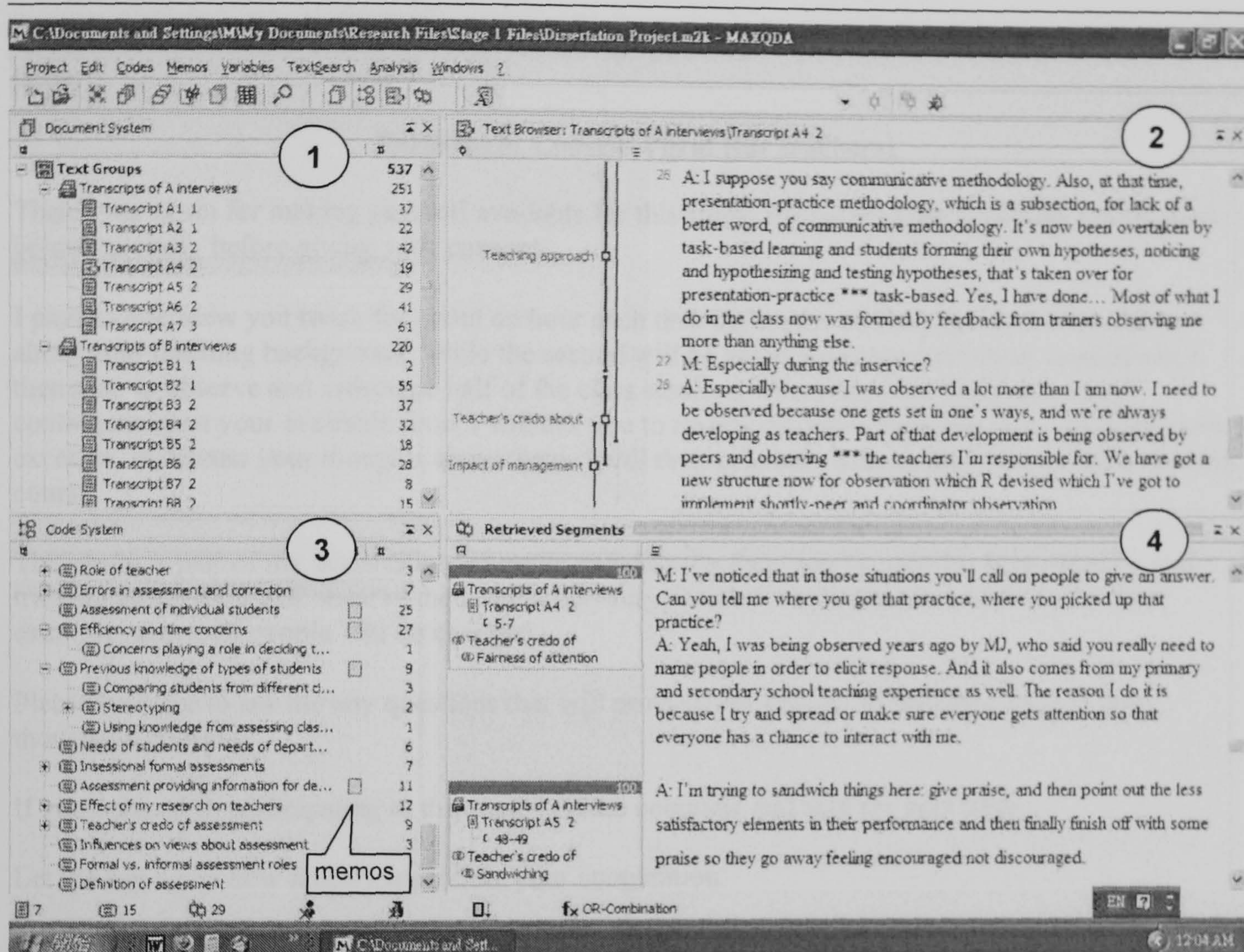
Issues to reflect on:

One key issue, which resonates with my own experience, is how the "tightening-up" of the syllabus and the assessments as a result of the university's review processes is influencing the teacher's work. Here's some analysis: it seems the review process requires clarification ("tightening-up") (as we discussed afterwards, B said democracy goes hand in hand with bureaucracy—the greater amount of information, the need to manage info, etc.) of aspects like the syllabus content and the assessment procedures, and this has led to significant time pressure to get a lot of content into a short time. This in turn means class activities are on a quick pace; (conjecture here: this quicker pace may influence the type or kind of assessment information teachers take in).

B doing the work on explicating his own implicit criteria, partly spurred on by my questioning, highlights the agency of the teacher ... and how researchers and teachers are in a mutually influencing relationship. How could I as a researcher try to prevent what B did? There's no ethical or practical way, and why should I need to?

Appendix 5 Using MAXqda

Below is a screenshot from MAXqda, with a brief description of how the software was used (see also 3.5.2).



The transcripts, typed in Microsoft Word and saved as Enriched Text Files, were loaded into MAXqda's document system (1). I then used the text browser (2) and went through each transcript to code segments (see analysis practice 4 under 3.5.2), based upon and also contributing to the codelist or code system (3); the indicated squares in the code system window stand for memos that I wrote to myself regarding concepts that were in development or questions that were raised by a segment of data or a group of segments. After coding one or more transcripts, I could retrieve and look at all the segments that were given a particular code or codes in the retrieved segments window (4), thus facilitating the analysis process (see 3.5.2, 3.5.3, and 5.3.1). In addition, clicking on a segment's accompanying box on the left in (4) brought up into the text browser (2) the transcript from which the segment was drawn; as Kelle (1995) points out in regard to computer-aided qualitative data analysis generally, this allows the analyst to look at the segment's context, so as to reduce the danger of fragmenting the data into unrelated parts.

Appendix 6 Consent forms used

The consent forms used in this study are provided below (see 3.6).

Consent form for teachers in Stage 1

page 1

Participant Consent Form [for teachers]

Thank you again for making yourself available for this study. Please read the following information below carefully before giving your consent.

I plan to interview you twice for about an hour each time before the course. The first time will be about your teaching background, while the second will be about your thoughts about assessment. I then plan to observe and videotape half of the class sessions. Then, after the observation, and contingent upon your available time, I will ask you to review activities from the class, including video excerpts, to discuss your thoughts about them. I will then conclude with an interview at the end of the course.

This study is part of my dissertation research, and thus the data may be used for both research and training purposes. I will observe the usual anonymity practices in publications and reports. For example, names of people will be changed.

Please feel free to ask me any questions that will provide you enough information to give your thoughtful consent.

If you consent to participating in this study, please complete and sign the next page.

Let me say again how much I appreciate your cooperation.

If you have any questions or concerns later, please contact me at:

Muchun Yin, Graduate School of Education, University of Bristol, 8/10 Berkeley
Square, Bristol BS8 1JA
muchun.yin@bris.ac.uk
0117 928 7175 (W)
0117 973 5459 (H)

page 2

Participant Consent Form

I would like to have a written record of your consent, so please tick the boxes below and sign and date below.

☐

I consent to being interviewed, recorded, and observed by Muchun Yin.

☐

I consent to such data being analyzed for research and training purposes and understand that as far as possible anonymity will be preserved if extracts are included in research publications or reports.

NAME:

SIGNATURE:

DATE:

Consent form for students in Stage 1 (initial)

This initial consent form was given to students who were present at the start of my Stage 1 class observations. Questions arising from this led me to revise it (see next version).

Student Consent Form

I am a doctoral student in the University of Bristol Graduate School of Education, and I am doing research into teaching. Your teacher has agreed to participate in this study and will be the focus of my research.

However, as part of my study, I will observe and videotape the classroom, and I will discuss my observations and videotape with the teacher. I would like to ask your consent to being videotaped and perhaps discussed.

The data may be used for both research and training purposes. I will observe the usual anonymity practices in publications and reports. For example, names of people will be changed.

Please feel free to ask me any questions that will provide you enough information to give your thoughtful consent.

If you consent to being videotaped and perhaps discussed in this study, please complete and sign the bottom of this sheet and return it to me.

Let me say again how much I appreciate your cooperation.

If you have any questions or concerns later, please contact me at:

Muchun Yin, Graduate School of Education, University of Bristol, 8/10 Berkeley
Square, Bristol BS8 1JA
muchun.yin@bris.ac.uk
0117 928 7175 (W)
0117 973 5459 (H)

Student Consent Form

I would like to have a written record of your consent, so please tick the boxes below and sign and date below.

☐ I consent to being recorded and discussed by Muchun Yin and the teacher in relation to this course.

☐ I consent to such data being analysed for research and training purposes and understand that as far as possible anonymity will be preserved if extracts are included in their research publications or reports.

NAME:

SIGNATURE:

DATE:

Consent form for students in Stage 1 (revised)

This is a revised version of the initial form I used; among other changes, this form makes clearer my use of the data, and it makes clear that there is a choice in the consent form options.

Student Consent Form

I am a doctoral student in the University of Bristol Graduate School of Education, and I am doing research into teaching. Your teacher has agreed to participate in this study and will be the focus of my research.

As part of my study, I will observe and videotape the classroom, and I will discuss my observations and videotape with the teacher. I would like to ask your consent to being videotaped and perhaps discussed by the teacher and myself.

The data will be used for research purposes only. I will observe the usual anonymity practices in publications and reports. For example, names of people will be changed. If I use videotape for purposes other than research, I will ask for your consent.

Please feel free to ask me any questions—about who I am, what I am doing, or anything else—that will provide you enough information to give your thoughtful consent.

The rules of the university require that I have a paper record of your consent. Please check the appropriate box below, complete the rest of the form, and return it to me.

Let me say again how much I appreciate your cooperation.

If you have any questions or concerns later, please contact me at:

Muchun Yin, Graduate School of Education, University of Bristol, 8'10 Berkeley
Square, Bristol BS8 1JA
muchun.yin@bris.ac.uk
0117 928 7175 (W) 0117 973 5459 (H)

Student Consent Form

Please tick the appropriate box below and sign and date below.

☐ I consent to being recorded by Muchun Yin in relation to this course, and I consent to such data being analysed and used for research purposes. I understand that as far as possible, anonymity will be preserved if extracts are included in research publications or reports. (If possible, please provide an e-mail account so that I can contact you in the future if necessary: _____)

☐ I do not consent to being recorded and discussed by Muchun Yin and the teacher in relation to this course.

NAME:

SIGNATURE:

DATE:

Consent form for Stage 2 focus group participants

The participants received two pages. The first (I) is the consent form. The second (II) is a questionnaire asking for general background information about the participant.

I. Interviewee Consent Form

Thank you again for making yourself available for this interview. Following guidelines for good research practice*, I would like to formally ask for your consent to participate in this study. Please read the following information below carefully before giving your consent at the end of the page.

I plan to interview you for about an hour about issues related to your assessment of students in language classrooms; this will be recorded on audiotape. I will use this data and the background data you provide in the next section to inform my dissertation research on teacher thinking in relation to assessment.

I will observe the usual anonymity practices (e.g., names of people will be changed in the transcript) and try my best to maintain participant confidentiality.

I will primarily use the data for research purposes. However, I may use data for presentation in the future (e.g., in a conference or publications); if so, I will still maintain informant anonymity.

Let me say again how much I appreciate your cooperation. Feel free to contact me if you are interested in the results of my analysis.

If you have any questions or concerns, please contact me at:

Muchun Yin
Graduate School of Education, University of Bristol, 8/10 Berkeley Square, Bristol BS8 1JA
muchun.yin@bris.ac.uk
0117 973 5459

* For further details, see the British Association for Applied Linguistics recommendations on good practice at <http://www.baal.org.uk/goodprac.htm>

To indicate your consent, please fill out the information below and tick the boxes.

Name (please print):

Signature:

Date:

- ☐ I consent to being interviewed and recorded by Muchun Yin.

- ☐ I consent to the data I provide being analyzed for research and presentation purposes, and understand that as far as possible anonymity will be preserved whenever data is presented.

II. Biographical Data

- 1. Name:
- 2. Nationality:
- 3. How many years have you taught English language to non-native speakers? How many of those years have you taught English for Academic Purposes specifically?
- 4. Please give a brief history of your English language teaching experience (including any teacher training qualifications); an example is given. If you have taught other subjects, please give them also.

1998 Completed CELTA course at University of Bristol
1999-2002 Taught general English at private language school in Barcelona, Spain. Students consisted mainly of Spanish children and youths.
2002-present Taught adult ESOL at a college in Manchester, UK. Students consist mainly of Chinese immigrants. Also worked as a manager from 2004 to present.

- 5. Would you describe yourself as a native speaker of English? If not, what is/are your native language(s)?
- 6. Besides your native language(s), what other languages can you use, at about what level of proficiency?
- 7. Have you ever had training specifically in regards to classroom assessment (this can include pre-service or in-service training)? If so, please describe briefly.

Appendix 7 Stage 1 class syllabus and scheme of work

The syllabus and scheme of work for the in-session class that the Stage 1 instructors taught are given below (see 4.2).

In-session EAP core syllabus

Objectives

Learners will develop their ability to:

- Participate effectively in seminars and tutorials
- Prepare and deliver an effective academic presentation
- Understand the structure and development of academic talks
- Produce an appropriate, coherent and cohesive academic written text
- Integrate source material into writing following appropriate academic conventions
- Take effective notes from written or spoken texts
- Read with greater efficiency and effectiveness
- Understand their learning needs and develop appropriate independent learning strategies and programmes

Syllabus

Speaking:

- Exchanging and reformulating factual information
- Expressing and supporting opinions
- Agreeing and disagreeing
- Turn taking (interrupting, bringing people in, holding the floor)
- Planning an academic talk
- Preparing and using visual aids
- Using transition and signalling devices
- Giving an academic talk
- Asking and responding to questions

Listening:

- Understanding a speaker's point of view and the development of an argument
- Recognising transition and signalling devices
- Understanding unknown words and phrases from context
- Listening for specific information
- Listening and note-taking

Writing:

- Understanding the structural elements of academic text-types (short essay, long essay, dissertation, research report, lab report)
- Understanding and using logical connectives in academic writing (comparison and contrast, exemplification, cause and effect)
- Summarising and paraphrasing from written sources
- Incorporating written and visual data into own writing using appropriate academic conventions

Reading:

- Understanding a writer's point of view and the development of an argument
- Recognising transition and signalling devices
- Understanding unknown words and phrases from context
- Reading for specific information
- Reading efficiently
- Reading and note-taking

Independent learning:

- Identifying learning needs and priorities
- Developing appropriate individual learning strategies and programmes

19th January – 19th March 2004: Scheme of Work

Speaking	Listening	Writing	Reading	Independent Learning	Negotiated Syllabus Element (non-EAP)
Focus Exchanging and reformulating factual information. Suggested Materials 1. SS pp.8-9 2. EASSS p.4	Focus Diagnostic Listening Test Suggested Materials SL p. 71	Focus Understanding the structural elements of academic text types. Suggested materials	Focus Diagnostic Reading Test Suggested Materials IELTS Specimen Materials 2003 pp. 17-20	Focus Needs analysis Suggested Materials 1. LC doc. 2. EASP pp. 107- 109 3. LLE pp 10 & 109, p.6	
Focus Using transition and signalling devices/ Giving an academic talk Suggested Materials 1. SS pp. 9-11 2. EASSS pp. 24-26	Focus Recognising transition and signalling devices. Suggested Materials 1. SL pp 53-57	Focus Logical connectives 1: Comparison and Contrast Suggested Materials 1. Approaches pp. 33-36 2. SW p.46 3. WAE p.141-149	Focus Understanding writer's view and development Suggested materials 1. EASR pp 6-10 2.	Focus Introduction to SAF Suggested Materials LLE pp. 20 - 23	
Focus Agreeing and Disagreeing Suggested Materials 1. SS p.20 2. STIE pp.84 - 86	Focus Recognising purpose and structure of lectures. Suggested Materials 1. EASL pp.15-17 2. SL p.72	Focus Logical Connectives 2: Exemplification Suggested Materials. 1. Approaches p.28 2. AWC p.29	Focus Reading and note-making 1 Suggested Materials 1. SSAW Unit 2 2. STIE p.62	Focus Discussion: Using English outside the classroom	
Focus Preparing and using visual aids Suggested Materials 1. SS pp 18-19 2. EASS pp. 10 -11	Focus Meaning from context Suggested Materials 1. EAS pp. 43-44 2. SL p.73	Focus Logical Connectives 3: Cause and Effect Suggested Materials 1. Approaches p.58 2. AWC p.58 3. WAE pp.130 - 140	Focus Understanding unknown words and phrases from context Suggested Materials 1. STIE pp. 76 - 77 2. SR pp. 38 -39	Focus Extending Vocabulary Suggested Materials LLE pp 27 - 43	
Focus Expressing and supporting opinions. Suggested materials 1. SS. p.24 2. EASS pp. 54- 56	Focus Listening for specific information Suggested materials 1. SL p.74 2. ALE pp5-6	Focus Summarising and paraphrasing 1 Suggested Materials 1. SSAW pp. 39 - 55 2. AWGS pp. 105 - 129 3. EASW pp. 46-48	Focus Reading efficiently Suggested Materials 1. SSAW pp. 15 - 17 2. STIE pp. 61 & 76	Focus Progress Check: Using the SAF	
Focus Asking and Responding to questions Suggested Materials 1. SS pp 24-26 2. EASS pp 47-48	Focus Listening and Note-taking Suggested Materials 1. SL pp. 51-52 2. EASL pp. 33 et seq.	Focus Summarising and Paraphrasing 2 Suggested Materials 1. Approaches p.129	Focus Recognising transition and signalling devices Suggested Materials 1. SR pp 45-50 2. EASR pp. 79-80	Focus Progress Check : Using English outside the classroom	
Focus Turn-taking Suggested Materials 1. STIE pp 86-88 2. EASS p.40	Focus Guessing vocabulary from context. Suggested Materials 1. ALE pp 9-11	Focus Quoting / Data Commentary 1 Suggested Materials 1. AWGS pp. 77 - 85	Focus Reading and Note-making 2 Suggested Materials 1. SR pp 56 - 57	Focus Progress Check: Extending Vocabulary	
Focus Seminar talks and Discussions Suggested Materials 1. SS pp 47-49 2. EASS pp 58-61	Focus Understanding meaning from context. Suggested Materials 1. ALE pp 120 - 123	Focus Quoting/ Data Commentary 2 Suggested Materials 1. WAE p.82	Focus Integrated Skills: Reading – note-making- summarising	Focus Self-assessment of progress made Suggested Materials: LLE	
Focus Giving an academic talk Group presentations and feedback this week	Focus Review Suggested Materials EALS pp 68 - 75	Focus Revision / Collaborative Writing task / Teacher Reformulation	Focus Integrated Skills : Reading – note-making - summarising	Focus Self-assessment of progress made Suggested Materials LLE	

List of Abbreviations:

EASS = English for Academic Study: Speaking
 EASL = English for Academic Study: Listening
 EASR = English for Academic Study: Reading
 EASW = English for Academic Study: Writing
 AWC = Academic Writing Course
 LLE = Learning to Learn English

SL = Study Listening
 SR = Study Reading
 SS = Study Speaking
 Approaches = Approaches to Academic Reading and Writing
 WAE = Writing Academic English
 SSAW = Study Skills for Academic Writing

SSH = A Study Skills Handbook
 AWGS = Academic Writing for Graduate Students
 LC doc. = Language Centre document
 ALE = Academic Listening Encounters

Appendix 8 Examples of teacher questioning in Stage 1

Provided below are excerpts of transcripts of classroom discourse from Stage 1; they illustrate the questioning patterns of CTA and CTB (see 4.4). Note: *** is unclear speech, text in brackets are notes, and underlined text are teacher questions (not including tag questions).

Excerpt of CTA's class on 11 February 2004

In this session CTA is discussing academic writing and has just given his students a handout called "Moral Standards and Social Organization." He is standing in front of the whole class.

Starting from 15:18

CTA: Before you read a text, think about what you already know about the subject, such as with an anthropological text. From your knowledge of the world, can you tell me what anthropologists study? [silence] SY, any ideas?

SY: No.

CTA: Can anybody help?

SK: Societies.

CTA: Yes. Human societies. Do they study large, industrialized, complex societies or small-scale simple societies? [silence] Normally, small societies. Sociologists study large-scale societies, cultural anthropologists study small-scale societies. Look at the text "Moral Standards and Social Organization". What do you expect the text to be about? [pause] "Moral Standards and Social Organization." In simple English, what's a moral standard, SE?

SE: Moral standards.... um... I can't say the word, but something...

CTA: Can you give me an example—[someone says something]—rules, rules of behavior, yeah. So moral means what's right and what's wrong. So rules of behavior of what's right and what's wrong. And social organization, what's the connection between rules of behavior and social organization?

SK: How the rules influence the society and how the society influences the rules.

CTA: Perfect. Perfect answer. Is this your subject?

SK: Now.

CTA: Well, you are answering well so far, SK. [steps to whiteboard] Any words or phrases you would expect to find in the text? ... [students say out words and CTA writes them on the board] "tradition," "religion," ... SA, what's the writer describing?

SA: The writer is describing *** and then he describes ***.

CTA: His ***?

SA: ***

CTA: What are the moral standards referred to in the title there, please, SY?

SY: ***

CTA: That's right, yeah. And what is the writer comparing? SP.

SP: ***

CTA: Right, well, just look through it very quickly for information. What is the writer comparing ***?

SP: Maybe behaviors of the people in his country and people in another country

CTA: Yes, exactly. ... You haven't studied the text carefully, but you were able to get information by looking through it quickly. That's exactly what I wanted you to do, so 'well-done.'

Excerpt of CTB's class on 10 March 2004

For this class session, CTB has students read through their countries' entries in *The World Guide* and has students try to detect bias and guess the authors. He then has them compare each others' countries. In excerpt 1, CTB has divided the class into groups of 2-3 and has been going around to each group to observe, facilitate, and contribute to their discussions; in that excerpt he is with the group consisting of SA, SM, and ST. In excerpt 2, CTB is at the front of the class and has gotten everybody's attention for a whole-class discussion.

Excerpt 1—starting from 17:36

CTB: We have to look for statements that seem to be positive, approving, and statements that are critical, disapproving.

SA: Here... [reading his text] "'reform' the Communist Party."

CTB: Yeah, there you go. That's ***, isn't it? What do we do when we put words in quotation marks? [unclear whether spoken to the group or to SA]

SA: Because you don't really believe it was a reform.

CTB: Yeah, it is, you're saying, "this is the word they're using, well, don't take it so seriously, let's think about it."

A: [expresses agreement]

CTB: Then that would be a point there. What kind of bias does that indicate?

SA: ***

SM: *** somewhere else it states ***

SA: I think ***

CTB: Even if you can't clearly identify what you think the writer means there, you could query that, you could say, "why don't you write it this way?"

SM: Yeah, yeah.

CTB: [To SA] How would you write it? I mean, would you just give it its name and not ***?

SA: Yeah. ***

Excerpt 2—starting from 18:06

CTB: We have another activity to do, so we'll wind this one up now. I just want to see where you think ... [gestures to SM] you said that you thought it's a very English point of view or, like, maybe British, maybe it's produced in Britain with a team of experts... fairly central point of view...

SM: No, left of center.

CTB: Left of center [gestures to students' left]... so not Tony Blair, then [some student chuckles], because he's over there somewhere now [gestures to students' far right]. [gestures to SJ and SY] What did you think? Did you think it was likely to be produced by a British team?

SJ: It's a bit undecided *** but for the moment ***.

CTB: But you feel there's some input from Spanish people?

SJ: Yeah.

CTB: *** consultation with Spanish historians and Spanish points of view.

SJ: Yeah.

CTB: But again center left?

SJ: Yeah.

CTB: It's very pro-international cooperation and very anti-acts of aggression, isn't it? It's very pro-open systems of government and very anti-totalitarian or extreme systems of government, yeah. [students nod in agreement] And of course not all histories are like that. Histories can be written in many, many ways. [gestures to SN and SC] What about you two? What did you feel? Did you think it could reflect a British point of view?

SN: More American...

CTB: More American, yeah... because of course the English language, it could be an American perspective.

Appendix 9 Interview guide for Stage 2 focus groups

The list of questions below served to guide my questioning in the focus groups. I did not ask these questions in this order or in the way they are written here; rather, I used them as a kind of checklist to make sure I covered the main areas I wanted to discuss. The parts in gray are meant to show how these interview questions “operationalized” the research questions (see 5.2.4).

Focus Group Interview Questions

1. How do you define “classroom assessment”? [As in Stage 1, this was to check teacher and researcher definitions.]

2. In my previous research, I worked with two teachers of English for Academic Purposes. Their practices were rather different, as described below:

Teacher 1 often called on students in front of the whole class and asked them questions that usually had a right or wrong answer (e.g., “What does a biologist do?”). To a lesser extent than Teacher 2, Teacher 1 also set up groups to do group work and went around observing them and occasionally giving oral feedback.

Teacher 2 often set up group work and then spent most of the class time observing and giving feedback to each group. Also, in front of the whole class, the teacher asked questions that often did not have a right or wrong answer, and anyone in the class could volunteer an answer.

Would you describe yourself as similar to either one of the teachers? If so, which one? If not, how would you describe your practices in regards to questioning and observing students? [This and the next question were to answer RQ6 about the impact of managerial action on assessment practices. I stopped using this question after the pilot because of reservations about its usefulness and time restraints.]

3. The two teachers were teaching a course in which there had been a change to a) a more explicit/specified syllabus, and b) uniform end-of-course summative assessments (that is, teachers teaching the same course had to conduct the same summative assessment tasks). One of the teachers felt a lot of time pressure and felt somewhat unable to do a lot of things he wanted to do in his class because of these changes. The other teacher felt positive about the changes.

Have you had a similar experience where the syllabus or summative assessments were changed like this? If so, please describe it briefly. How did you feel about the change(s)? Did your teaching practices change as a result?

4. While both of these teachers were observing their students, they found that an initial judgment that they had made was incorrect and had to be changed, as the example below shows:

I asked one teacher why he had grouped the students for an activity the way he had. He replied, “Basically the reason that I chose SY and SI from Iran and Japan to share the task of presenting one of the texts to SJ, is that these three students were noticeably in the first part of the lesson of a slightly lower level in communicative ability and in extracting information from the texts. That was my initial [assessment]. ... I made that snap decision, but subsequently found that SJ in the discussion was actually more eloquent and had a better ability to draw on lexis than SE, who was sitting here.”

Have you ever had a similar experience of realizing that an impression that you had of a student’s language ability was actually incorrect? If so, please give an example: in particular,

how did you know that your previous impression had been inaccurate? [This and the next question were to answer RQ5 about improving the quality of teachers' impressionistic knowledge of students' language abilities.]

5. How do you know that your impressions of a student's language abilities are correct?

6. When talking about what they were thinking as they observed a student's language performance in class, the teachers sometimes said they would imagine the student in a particular situation (like having a conversation with British classmates, or performing on the end-of-course summative assessment) and make judgments based on that. Have you had similar thinking when observing students? [This and the next question were to answer RQ7 about projection and stereotyping as aspects of assessment thinking.]

7. The teachers occasionally referred to general preconceived ideas of what a student or group of students would be like, often based on experience with similar current and past learner groups. These preconceived ideas could be based upon ethnicity, nationality, or language, or upon course groups, and they seemed to influence the teachers' assessment of students, especially the feedback they gave. Here are two examples:

Teacher: "I'm tuning in to the two [students] on my right, making individual notes about features of their pronunciation, their level of fluency, and for example Student C I would be writing down that she speaks in a very slow and staccato fashion. Her language is quite accurate but she needs to improve her fluency. And like a Chinese speaker, she also pronounces every syllable equally when she speaks English."

Teacher: "So he then demonstrated quite a rapid fluency. I also wrote something down, quite a structural fluency in his argument. ... Certainly with Europeans, like Spanish and Italian students, they have a verbal, an oral confidence, a willingness to express themselves in spoken language, which makes them fairly fluent, with lots of errors, whereas [the] Chinese education system, Japanese education system will tend to restrain—and maybe there are cultural factors as well—restrain their willingness to experiment and make mistakes, so that it's interesting that I noticed he was good when he was speaking freely, and noticed he had problems when he was analyzing text."

Have you used this kind of preconceived idea in your own thinking when you assess students? If so, please give an example.

Appendix 10 Triangulation of Stage 2 coding

This appendix gives some detail of how the Stage 2 coding was checked by my colleague C (see 5.3.1). First is the introductory e-mail setting out my request to my colleague. This is followed by the brief codelist I gave to C. Finally, an excerpt from the transcript I gave to C is provided, along with the coding by C and by myself (note: a few of my codes are more specific than C's; I did not put these more specific codes in C's codelist).

Introductory e-mail

Hi C,

Thank you very much for being willing to help me out with doing a coding check.

Let me give you some background on this. In Stage 2 of my study, I conducted focus groups with teachers at a couple university language centres. I had some initial questions and data prompts (from my Stage 1 data) that I wanted to ask the teachers to respond to, but I also asked follow-up questions. One of the attachments is a section of the transcript from one of these focus groups.

What I would like you to do is:

- 1) read over the whole transcript briefly to get a basic feel for it;
- 2) look over the coding information (the second attachment); the codes I've given you are not exhaustive but are the main areas I've been looking at in my analysis.
- 3) using the codes I've given you, go through the transcript and mark any parts you feel fit particular codes. (By the way, one bit of transcript can have multiple codes.) I don't know whether it's more convenient for you to print it out and code it, or perhaps you can use WinMAX to code it (if you have it) in which case you'll have to save the RTF document as plain text, I think. The main thing is that I can know which parts you've labeled under which coding categories.

Let me know if there's anything unclear. If it's all OK, then when you finish it could you leave it with the porter in 35 so I can pick it up?

I hope this will not be too much for you; I mentioned 30-60 minutes on Wednesday but as I look at it I'm not sure, since I realized that you'd need quite a bit of transcript text to work on if the coding was going to be of any use (I think it would be pointless to have you code 2-3 categories in 2-3 pages of text.). If you find it burdensome, please let me know and I can cut it down somewhat, though.

OK, thanks again!

Muchun

Code list for C

Hi C,

I hope this chart is relatively clear to you. There are three general topics (in grey) with sub-topics below them. When you code the transcript section, if you come across something that falls under a specific code (like “sources of information about students”), then you can give it that label; if it’s related to the more general topic but doesn’t fall under one of the sub-topics, then use the general label (e.g., “quality of teacher assessments”).

The labels are only suggestions; you can use whatever label you wish, as long as I know which code it is.

Muchun

Possible label	Code	Meaning
TA quality	Quality of teacher assessments	Anything that relates to the quality of teachers’ assessments of students
Sources	Sources of information about students	What are the sources from which teachers draw info about student ability?
Receptive	Weakness in assessing receptive skills	What are potential “blindspots” when teachers assess students’ listening and reading ability informally?
Ambiguity	Ambiguity of silence	How do teachers interpret student silence?
Not knowing	Students not knowing they are being assessed	Related to how students may not always be aware they are being assessed by the teacher
Stereotyping	Stereotyping	I found in the case studies that teachers often had “stereotypical” knowledge about students. Anything related to this.
What stereo.	What is stereotyped	What does that stereotyped knowledge consist of?
Dangers stereo	Dangers	What are problems with stereotyping?
Good stereo.	Not negative	What are benefits of stereotyping?
Look individual	Useful info but still look at individual	Stereotypes are helpful but teachers must still look at the individual
Example stereo.	Examples of ethnic/nationality stereotypes	Examples of stereotypes
Projection	Projection	In the case studies, I found that teachers often imagined or “projected” a student performance into a target language use (TLU) situation (here, academic situations on the students’ courses). Anything related to this.
Counterexample proj.	Counter examples	An example of a teacher who did not feel he used projection
General proj.	General sense only/not subject specific	These projections are of a general academic nature, not specific to particular subjects like business or engineering.
TLU knowledge	Knowledge of future academic situation	What and how do teachers know about the students’ future academic situation?

Excerpt of transcript sent to C, with my and C's coding

C's coding	Transcript excerpt	My coding
	Participants: SL (manager/teacher), CC (teacher), SM (teacher), and Muchun Yin (MY)	
	1 MY: [Introductory comments] How do you find out about students' language ability in your classes?	
	2 SL: How I do it.	
	3 MY: Yeah, that's a general question.	
TA quality/ sources {	4 SL: How we do it is we have various formal assessment measures and also we use our experience and professional know-how to informally assess students as they're going about their daily business.	TA quality/ sources/ formal assess- ments {
	5 MY: What do you mean by daily business?	
	6 SL: In class.	TA quality/ sources/ interactions {
	7 MY: Right. You mean in terms of their interactions with you?	
TA quality/ sources {	8 SL: Well, yeah. In terms of their interactions with each other and with me and you know the progress they're clearly making during the course of a 9 or 10 week or 11 week program. That's a start. If you want to talk about the formal assessment measures we have, that's quite lengthy.	
	9 [SL discusses placement test; usually accurate but sometimes not; often just broad placement b/c only two or three class groups]	
Projection/ TLU knowledge {	10 CC: But we don't just assess students on formal tests. I mean it depends very much like what we're teaching students; if we're teaching a writing course, or if we're teaching a speaking/listening course then you're assessing different things, and if they're writing essays during the term, then you're looking at the essays and first drafts and second drafts or whatever, and you're comparing what they've come up with with some kind of idea of what good writers in the EAP context should be able to do with the information. So they should be able to organize it and develop and they should be able to make it coherent and they should be able to talk about quite abstract ideas and have the language that's necessary to express those abstractions. When you're doing writing, you're getting constant product from students, and yet you're constantly assessing and re-assessing as they work through those.	TA quality/ sources/ products { Projection/ TLU knowledge {
	11 MY: So you're actually getting quite a bit of information...	

C's coding	Transcript excerpt	My coding
TA quality/ sources	12 CC: A huge amount; a huge amount, I mean when there's product involved. That's true of speaking as well, because obviously with speaking, then during the course of a term we'd be witnesses to their speaking whether it's an individualized presentation that they've prepared or whether it's a non-prepared interaction with their peers, whatever. There is some ***[main product?].	TA quality/ sources
	13 SM: For me, there is a sort of difference in the way we treat the different skills. There's a distinction between the productive skills and the receptive skills in the sense that the productive skills you can—I believe—you can—reading and writing—you can be sort of assessing on the basis of coursework throughout the whole term	
TA quality/ sources	14 CC: Do you really mean reading or do you mean writing?	
	15 SM: I mean the productive skills.	
	16 CC: You said reading and writing.	
	17 SM: Did I?	
	18 CC: Yeah.	
	19 SM: Sorry. The productive skills writing and speaking.	
	20 CC: Clarification for the microphone.	
	21 SM: You can test on the basis of performance throughout the whole term. With listening and reading—I'm the listening and speaking coordinator—certainly for listening I have a bit of a problem about the continuous assessment of listening. What we have done in the past is that all the teachers who teach students—and it doesn't matter whether they're teaching a particular student reading or writing or listening and speaking—have given assessments of listening skills on the basis of how students respond to the teacher speaking or the other students speaking in class. I personally have a bit of a problem with that because I don't think it's the same kind of listening that we're really preparing them for, which is—the main kind of listening that we're preparing them for is listening to lectures and taking notes, and for me that's a different kind of listening.	TA quality/ receptive
	22 MY: So are you saying that—I just want to... let me get this straight—for you, you see an issue with the assessment of listening or of receptive skills generally. For example, with listening often it's seen as a student's ability to understand the teacher...	

C's coding	Transcript excerpt	My coding
TA quality/ receptive	23 SM: What I'm saying is if you assess listening on the basis of how students appear to be understanding what's going on in the classroom situation—and it could just be some group work or something like that—then I think that's not the same as actually listening to a say lecture.	TA quality/ receptive
	24 SL: I think the difference is really social listening as opposed to formal listening or something like that.	
TA quality/ receptive	25 CC: But you can get academic contexts in which teachers are talking to students and they're expecting students to respond to them, I mean whether that's like an interactive lecture or that's a tutorial or whether that's in the corridor or whatever. There are different kinds of listening but I think you may need actually to assess different kinds of listening as well, because they don't just listen to lectures.	TA quality/ receptive

